

Validation of automated post-adjustments of HDR prostate brachytherapy treatment plans by quantitative measures and oncologist observer study

Frida Dohlmar^{1,2,*}, Björn Morén³, Michael Sandborg^{1,2}, Örjan Smedby⁴,
Alexander Valdman⁶, Torbjörn Larsson³, Åsa Carlsson Tedgren^{1,2,5,6}

¹Medical Radiation Physics, Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

²Center for Medical Image Science and Visualization, CMIV, Linköping University, Linköping, Sweden

³Department of Mathematics, Linköping University, Linköping, Sweden

⁴Department of Biomedical Engineering and Health Systems, KTH Royal Institute of Technology, Stockholm, Sweden

⁵Department of Medical Radiation Physics and Nuclear Medicine, Karolinska University Hospital, Stockholm, Sweden

⁶Department of Oncology Pathology, Karolinska Institute, Stockholm, Sweden

ABSTRACT

PURPOSE: The aim was to evaluate a postprocessing optimization algorithm's ability to improve the spatial properties of a clinical treatment plan while preserving the target coverage and the dose to the organs at risk. The goal was to obtain a more homogenous treatment plan, minimizing the need for manual adjustments after inverse treatment planning.

MATERIALS AND METHODS: The study included 25 previously treated prostate cancer patients. The treatment plans were evaluated on dose-volume histogram parameters established clinical and quantitative measures of the high dose volumes. The volumes of the four largest hot spots were compared and complemented with a human observer study with visual grading by eight oncologists. Statistical analysis was done using ordinal logistic regression. Weighted kappa and Fleiss' kappa were used to evaluate intra- and interobserver reliability.

RESULTS: The quantitative analysis showed that there was no change in planning target volume (PTV) coverage and dose to the rectum. There were significant improvements for the adjusted treatment plan in: V150% and V200% for PTV, dose to urethra, conformal index, and dose nonhomogeneity ratio. The three largest hot spots for the adjusted treatment plan were significantly smaller compared to the clinical treatment plan. The observers preferred the adjusted treatment plan in 132 cases and the clinical in 83 cases. The observers preferred the adjusted treatment plan on homogeneity and organs at risk but preferred the clinical plan on PTV coverage.

CONCLUSIONS: Quantitative analysis showed that the postadjustment optimization tool could improve the spatial properties of the treatment plans while maintaining the target coverage. © 2022 The Authors. Published by Elsevier Inc. on behalf of American Brachytherapy Society. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Keywords:

Prostate brachytherapy; Automated treatment planning; Inverse treatment planning; Dosimetric indices; Observer study

Introduction

Treatment planning of prostate high-dose-rate (HDR) brachytherapy (BT) has historically been performed by manual adjustments of the dwell times and by using graphical tools to drag the isodose curves to the desired position (1). Manual treatment planning is time demanding and dependent on the skill of the treatment planner (2). Optimization algorithms – so called inverse treatment planning or inverse planning – are included in modern treatment plan-

Received 16 September 2022; received in revised form 9 December 2022; accepted 9 December 2022

Disclosure: This work was supported by grants from the Swedish Cancer Society (CAN 2018/622 and CAN 2017/1029), LiU Cancer, Region Östergötland (RÖ-966539) and the Swedish Research Council (VR-NT 2019–05416). There are no conflicts of interest.

* Corresponding author. Linköping University Hospital, O-huset Ingång 34 pl 08, S-58185 Linköping, Sweden.

E-mail address: frida.dohlmar@liu.se (F. Dohlmar).

ning systems (TPSs) for a more automatized approach (3). Inverse planning is frequently used for prostate and breast BT but is currently rarely used in cervical BT. Unfortunately, treatment plans created by inverse planning often need manual adjustments to be clinically accepted (4). For example, inverse treatment plans often have more varying dwell times, where fewer dwell positions are used than in manually created treatment plans (5). Fewer dwell positions necessitate longer dwell times, resulting in more heterogeneous irradiation of the target, with larger volumes irradiated to a level above the prescription dose. The dose volume histogram (DVH) is a 2D representation of the 3D dose distribution and does not yield information on the spatial locations of these overdosed volumes, which may be relatively spread out or more focused to form contiguously overdosed volumes. Lack of clinical evidence of the impact of overdosed volumes for prostate BT has been reported (6) but also indications on that the percentage of the prostate target volume treated to $\geq 200\%$ of the prescription dose influences urinary toxicity (7,8) and some argue that it would be reasonable to keep these volumes as small as possible (9,10). For breast BT there is evidence that the volume irradiated to 150% of prescription dose correlates with fat necrosis (11). Recommendations for breast BT (12) state the importance of inspecting and adjusting the 3D dose distribution for extensive 200% isodose volumes. Such adjustments are also, even if not explicitly stated in recommendations, common for prostate BT at many clinics. Furthermore, a more heterogeneous dwell time loading is less robust, meaning that a small shift of a catheter's placement, relative to the target and the organs at risk (OARs), can give larger changes in the dose delivered to the patient than would have been the case with a more homogeneous dwell time pattern (13).

An optimization tool developed to automatically perform the before-mentioned manual adjustments, resolving contiguously overdosed volumes, has been developed by Morén *et al.* (14). This optimization tool aims at improving the spatial properties of a tentative treatment plan while retaining (or improving) the DVH parameters for target coverage and the OAR sparing constant. The spatial properties that the optimization tool focuses on are the contiguously overdosed and the contiguously underdosed volumes. These volumes can be set at chosen levels and are henceforth denoted as hot and cold spots, respectively. Spatial properties of treatment plans are not possible to quantify by DVH parameters as there is no spatial information included in the 2D DVH. Hence, for the same DVH, the prevalence of hot spots can vary, and either be spread out or focused to fewer and larger volumes. Further, optimization algorithms that only optimize on the DVH parameters will not prioritize spreading out the high dose volumes. Many optimization algorithms in commercial TPSs include a dwell time deviation constraint which aims at producing more homogeneous dwell time patterns, and, indirectly, more homogeneous dose distributions. However, spatial properties

are not explicitly modelled and only indirectly affected by dwell time constraints. The optimization tool in Morén *et al.* (14) aims at minimizing the total hot and cold spot volumes. It was developed to replace the manual postprocessing after inverse planning. The time required to inspect and manually adjust the 3D dose distribution after inverse planning could be shortened and the result would be less dependent on skill of personnel.

The quality of treatment plans can be judged and compared quantitatively by DVH parameters and qualitatively by observer studies. The initial evaluation of the current postprocessing tool was performed with a few clinical plans (14) by comparing DVH parameters, available in clinical TPS, and complemented by using a novel quantitative analysis of the volume of contiguous hot and cold spots.

Human observer studies that compare treatment plans, or methods, against each other are quite rare in radiotherapy literature. In external beam radiotherapy (EBRT), Petersson *et al.* (15) and Kyrouti *et al.* (16) used observer studies based on pairwise comparisons to evaluate the performance of different treatment techniques and different treatment planning methods. Their data were analyzed using statistical methods for ordinal data, such as one-sided sign tests and Pearson's chi-squared test. In HDR BT, observer studies have been used to compare different treatment planning methods; one did not analyze the data with statistical methods (17), one did hypothesis testing but did not describe the method (18), and one used the paired two-sided Wilcoxon signed-rank test (19).

In diagnostic radiology, human observer studies are more common and used to compare the quality of images from different imaging modalities or image acquisition settings; this is referred to as visual grading. More precisely, visual grading is when observers (here radiologists) assess the visibility of specific anatomical structures, according to well-defined criteria (20) on a predefined ordinal scale. The assessment can be done pairwise, where the observer grades one of the images in comparison to another image, that is, the standard clinical protocol. Pairwise observer studies may be preferred to increase sensitivity of appreciating a small difference in image quality, if the difference between images is expected to be small (21). The data from visual grading studies can be analyzed using statistical methods for ordinal data, for example, ordinal logistic regression ("visual grading regression") (21).

In radiotherapy treatment planning, oncologists have access to objective quantitative DVH information, but still visually evaluate the 3D dose distribution registered on the patient anatomy. We argue that it is important to evaluate both subjective and objective information when assessing a dose distribution, and to adapt the analysis methodology to the type of data used in either case.

The aim of this study is to assess the performance of the postadjustment optimization tool on its capability to improve the spatial properties of clinically approved prostate

BT treatment plans. A quantitative comparison, using established DVH parameters and the aforementioned measures of volumes of hot spots is performed on a larger patient cohort than that used in a prior study (14). As a complement, an observer study with radiation oncologists is performed using visual grading statistical analysis tools (21), novel within radiation therapy. To ensure a thorough comparison, the RATING protocol (22) is applied.

Methods and materials

The study compares clinical treatment plans for prostate cancer with treatment plans optimized with a postadjustment tool. The two are denoted clinical and adjusted treatment plans, respectively.

Patients

The patients were treated with a combined treatment of 50 Gy in 25 fractions (fx) with EBRT and a HDR BT boost with 20 Gy in two fx (given 2 weeks apart: before, after, or in the middle of the EBRT) in accordance with the Swedish national care program (23). The study comprises 25 cases consecutively taken from previously treated patients at Linköping University Hospital; only the first fraction was used in this study. The patients had a mean planning target volume (PTV) of 40.4 cm³, ranging from 24.3 cm³ to 60.3 cm³. The treatment plans included on average 16 catheters, ranging from 11 to 19. The Swedish ethical review authority has given consent to use the anonymized patient data in the study (EPM 2020–03331).

Quantitative evaluation parameters

The treatment plans were evaluated using the DVH parameters proposed by GEC/ESTRO (24). For the PTV and the clinical target volume (CTV), these are the dose that 90% of the volume receives ($D_{90\%}$) and the volume that receives at least 100% of the prescribed dose ($V_{100\%}$). For the rectum, the dose to the hottest 2 cm³ (D_{2cc}) is used; for urethra, the dose to the hottest 0.1 cm³, 10% and 30% ($D_{0.1cc}$, $D_{10\%}$, $D_{30\%}$) are used. For PTV, the conformity index (COIN) (25), and the dose nonhomogeneity ratio (DNR) (26) are used. The total reference air kerma (TRAK) was taken from the TPS for all treatment plans.

In addition to these established measures available in the clinical TPS, the contiguous volumes of hot spots, set at level $\geq 200\%$ of the prescription dose, were compared. To calculate the hot spots the RT Dose-file was exported to MATLAB (MathWorks, Inc, Natick, Massachusetts), and the MATLAB function *bwconncomp* was used to calculate the number of dose points in each hot spot.

Clinical treatment planning

Clinical treatment plans were created in BrachyVision version 15.6 (Varian Medical Systems Inc, Palo Alto, California) for a GammaMed Plus ¹⁹²Ir source, by the medical physicist participating in the treatment. Transrectal ultrasound (TRUS) images of the patients in the lithotomy position were recorded using a Flex Focus 400 (BK Medical Holding Company, Inc, Burlington, MA), and a stepper (Civco Medical Instruments Co Inc, Coralville, Iowa) using Vitesse version 2.5 (Varian Medical Systems Inc, Palo Alto, California). A slice thickness of 1 mm was used. The images were imported to BrachyVision, where oncologists contoured the prostate – which here is equal to the CTV – and the urethra. The PTV is contoured automatically with a margin of 3 mm in all directions except the cranio-caudal, where no margin is added. The rectum is contoured as a 3 mm thick structure after the surface of the ultrasound (US) probe. The urethra is contoured as a circle with a diameter of 5 mm around the catheter in the urethra as seen on the US images.

Treatment planning was performed with the aim of giving the PTV 10 Gy, as the national recommendation is to treat the whole prostate gland. No information about biopsy results and clinical staging was used in this study. The needle placements were manually planned based on a needle matrix. The inverse optimization tool AVOL (27) was used for treatment planning, with a maximum dwell time of 11 s and a minimum of 0.3 s for a nominal source strength of 40.7 mGy/h. A template with objectives was used for the inverse treatment planning. Manual adjustments after the optimization are rare in the current clinical practice at Linköping University Hospital, but the objectives of the inverse planning method are sometimes modified to improve the treatment plan. In some cases, the planned needle placement could not be reached due to, for example, the pelvic bones covering the prostate.

Treatment plans were anonymized when exported from the TPS. All treatment plans were imported to a single fictive patient in BrachyVision version 16.1, to make the observer study more time effective. The treatment plans were recalculated with its TG43 dose calculation algorithm using a voxel size of 0.05 × 0.05 × 0.1 cm³ covering the whole image set. This was done to certify that all treatment plans were calculated using the same version of the algorithm, since the TPS underwent a software update between the treatment of the patients and the study. The treatment plans were named consecutively with a number from 1 to 25.

Optimization tool

The optimization tool (14) was applied to the clinical treatment plans in a research environment separate from clinical systems. The tool has also recently been implemented and evaluated in a research version of a clinical

TPS (28). In the objective function of the tool, there are penalties for each pair of PTV dose points in which the doses are both too high or too low. The penalties are scaled with the reciprocal quadratic Euclidean distances for the pair of points. The penalties for both too high and too low doses are based on sigmoid functions, which are smooth approximations of the Heaviside step function. The value of β , which controls the steepness of the slope of the sigmoid function, was set to 2 for both high and low doses. A higher value of β means that the sigmoid penalty function is steeper. The weights controlling the trade-offs between hot spots and cold spots were set to 1 and 0.6, respectively. Dose thresholds for the sigmoid functions were set to 20 Gy (200% of the prescription dose) and 9.8 Gy for high and low doses, respectively.

Constraints on dosimetric indices for PTV, CTV, and OAR, as well as maximal doses to OAR and normal tissue, were based on the values from the clinical plan. For the PTV, there were constraints on D_{90} , D_{95} and D_{99} , and for the CTV there was a constraint on D_{90} , while for the rectum and urethra there were constraints on the portion that received at most 6 Gy and 10.7 Gy, respectively. Sigmoid functions were used to formulate the DVH parameter constraints. Values of β were set to 4 for PTV, CTV and the rectum, and 8 for the urethra. All the beta values were kept from the previous study (14), in which they were tuned to reproduce good approximations of the DVH parameters in the clinical plans. The same dwell time constraints as for the clinical plan were used.

The tuning of the parameters mentioned above was done by optimizing 12 treatment plans. The prioritization of hot and cold spots was modified in a trial-and-error manner until the PTV and CTV coverage and the doses to the OARs were in the same range as in the clinical treatment plans, but with smaller hot and cold spots. The treatment plans were compared using DVH parameters and by visual inspection.

After postprocessing by the stand-alone optimization tool (14), the adjusted treatment plans were imported back to BrachyVision so that the final dose distributions were calculated in the same way as the clinical treatment plans.

Observer study

An observer study was set up using Plan Evaluation (Varian Medical Systems Inc, Palo Alto, California). Eight oncologists – here called “observers” – from six clinics and with at least 5 years of experience in treating HDR prostate BT participated in the study. Each patient’s case was placed in separate course named after its number. The adjusted and the clinical treatment plans were randomly named A and B and shown pairwise. The observers were instructed to choose which of the two treatment plans they would use to treat the patient with. No information about the patient and their medical history was included. The observers also graded treatment plan B compared to plan

A using three criteria: homogeneity (the distribution of hot and cold spots), PTV coverage, and normal tissue sparing (including OAR sparing). The detailed grading steps are seen in Table 1.

Before the study, a meeting was held with the participating observers to get a consensus regarding the interpretation of the criteria and the grading steps. The observers were given individual scoresheets stating the order in which the evaluations were to be performed. The order was randomly chosen for each observer. The observers were allowed to look at each patient case only once, and not go back and review a case a second time. The observations were for the majority of observers done through Zoom (Zoom Video Communications, Inc, San Jose, California). There was a possibility to comment on the treatment plans. The observers decided, after their own clinical practice, how they preferred the dose distributions to be shown (color wash, isodose curves or a mix), which dose levels to be shown and which DVH parameters to evaluate.

To evaluate the intraobserver reliability, two randomly selected treatment plans, number 22 and 6, were included twice as numbers 26 and 27. For each observer the repeated treatment plans were inserted so that there were 10 plans between the original and the copy.

A pilot study of 10 patient cases was performed before the main study to test the workflow of the observer study and the statistical analysis tools. The two observers in the pilot study were not participating in the main study.

Statistical analysis

The statistical analysis of the DVH parameters was performed in SPSS Statistics version 28 (IBM, Armonk, New York) using a paired *t*-test. The significance level α was set to 0.05.

In analysing the observer study a mixed-effects ordinal logistic regression model was defined with observer and patient identity as random effects (29). These statistical calculations were performed in R, version 4.0.2 (The R Foundation for Statistical Computing, Vienna, Austria) using the *clmm* command. The goodness of fit was reported using McFadden’s pseudo R^2 (30). Interobserver and intraobserver reliabilities were described with the Fleiss’ kappa (κ) and weighted kappa (κ_w), computed with SPSS.

Results

The quantitative information for the treatment plans is presented in Table 2. The adjusted treatment plans are equivalent to the clinical treatment plans on PTV coverage and rectal doses. There are significant changes in the volumes of high doses ($V_{150\%}$ and $V_{200\%}$ for PTV) urethra doses, with the adjusted treatment plans having smaller values of $V_{150\%}$ and $V_{200\%}$ and lower dose to the urethra. There are also significant differences in COIN and DNR,

Table 1
The grading steps for the observer study

Grading	Explanation
1	Confident that treatment plan B is better than treatment plan A.
2	Suspect that treatment plan B is better than treatment plan A.
3	The treatment plans are equivalent.
4	Suspect that treatment plan B is worse than treatment plan A.
5	Confident that treatment plan B is worse than treatment plan A.

Table 2

Mean values of dosimetric indices, for the size (number of voxels) of the four largest contiguous hot spots (doses above 200% of the prescription dose) and the total reference air kerma (TRAK) for the clinical treatment plans and the adjusted treatment plans. *t* test has been used for statistical analysis with a significance level α of 0.05. The adjusted treatment plan is equal =; is better +, is worse -. A \checkmark shows when the adjusted treatment plan fulfills what the post-adjustment aimed at

		Clinical treatment plan	Adjusted treatment plan		<i>p</i> -value
PTV	D90%	10.28 Gy \pm 0.42	10.30 Gy \pm 0.45	= \checkmark	0.371
	V100%	92.00% \pm 3.00	91.98% \pm 2.87	= \checkmark	0.711
	V150%	29.73% \pm 2.52	27.61% \pm 3.05	- \checkmark	< 0.001
	V200%	10.78% \pm 1.24	9.55% \pm 1.23	- \checkmark	< 0.001
CTV	D90%	10.96 Gy \pm 0.17	10.99 Gy \pm 0.18	= \checkmark	0.085
	V100%	96.80% \pm 1.34	96.59% \pm 1.56	-	0.011
Rectum	D2cc	3.52 Gy \pm 0.50	3.50 Gy \pm 0.49	= \checkmark	0.366
Urethra	D0.1cc	10.92 Gy \pm 0.07	10.86 Gy \pm 0.07	-	< 0.001
	D10%	10.96 Gy \pm 0.08	10.88 Gy \pm 0.08	-	< 0.001
	D30%	10.80 Gy \pm 0.07	10.80 Gy \pm 0.06	= \checkmark	0.257
COIN		0.76 \pm 0.03	0.78 \pm 0.03	+ \checkmark	< 0.001
DNR		0.32 \pm 0.03	0.30 \pm 0.04	+ \checkmark	< 0.001
TRAK (mGy at 1 m)		0.32 \pm 0.03	0.30 \pm 0.04	+ \checkmark	< 0.001
Hot spots	largest	4145 \pm 2088	2674 \pm 1098	+ \checkmark	< 0.001
	Second largest	3307 \pm 1475	2146 \pm 768	+ \checkmark	< 0.001
	Third largest	2120 \pm 795	1622 \pm 537	+ \checkmark	< 0.001
	Fourth largest	1585 \pm 647	1378 \pm 432	=	0.065

with the adjusted treatment plans having on average better conformity (a higher COIN) and better homogeneity (a lower DNR). The TRAK for the adjusted treatment plans was on average significantly lower than for the clinical plans.

From the analysis of the four largest hot spots, it can be seen that the three largest ones are significantly smaller in the adjusted plans. Figure 1 shows the data from Table 2 as box plots of the relative difference between the clinical and adjusted treatment plans; the largest changes are clearly in the high dose regions (see $V_{150\%}$, $V_{200\%}$ and volumes of largest hotspots). Figure 2 is a cumulative histogram that shows how the hot spots of all patient cases have been reduced in size for the adjusted plans. For example, the number of hot spots that contains more than 3000 dose points are 35 for the clinical treatment plans and 12 for the adjusted plans.

Figure 3 shows an example of the dose distribution and the DVH that the observers graded. Seven of the observers completed all the gradings that were asked for. One observer made all the observations but did not answer all the questions. The analysis was done on all the observations, ending up with 215 grades for preferred plan, 213 for homogeneity, and 189 for PTV coverage and normal tissue sparing. The observer study showed that the oncologists

preferred the adjusted treatment plan in 132 cases and the clinical plan in 83 cases, including the doubly observed cases (one patient observation was not done). The two local oncologists chose the adjusted plan in 13 cases and the clinical plan in 41 cases, while the other six oncologists chose the adjusted and the clinical plan in 118 and 41 cases, respectively. Two observers stood out: one chose the adjusted treatment plan for only three patients, and one chose the clinical treatment plan for only 2 patients. There was only one patient case for which the adjusted treatment plan was chosen by three observers; for all other patients the adjusted treatment plan was chosen by four or more observers. There was one case where all observers preferred the adjusted treatment plan, and two where all but one observer preferred the adjusted plan.

Two of the treatment plans were judged twice by all the observers. In 13 of the 16 double observations the same treatment plan was chosen. The intraobserver reliability ranged from 0.44 to 1.0, meaning moderate to almost perfect agreement. The median weighted kappa was 0.84, meaning an almost perfect agreement. The interobserver reliability analysis showed no agreement (kappa < 0) to slight agreement (kappa 0–0.20).

The gradings were also resorted to represent whether the observers favored the adjusted or the clinical treatment

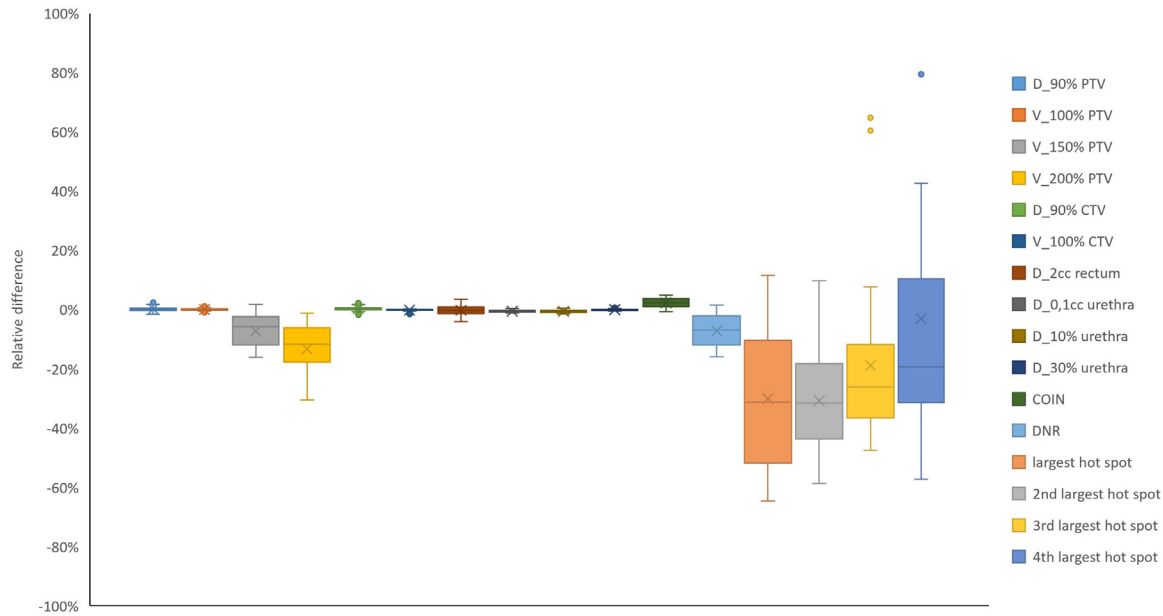


Fig. 1. Boxplot for relative difference between the clinical treatment plans and the adjusted treatment plans for the dose volume histogram parameters and the four largest contiguous hot spots. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 3

Mixed-effect ordinal regression results based on 216 comparisons of clinical and adjusted treatment plans. A positive regression coefficient meaning the adjusted treatment plan is preferred

Criterion	Regression coefficient (b) for planning method \pm SE	e ^a	p-value	McFadden pseudo R ²
Preferred planning method (adjusted vs. clinical)	0.474 \pm 0.145	1.61	0.0011	0.037
Homogeneity	0.611 \pm 0.139	1.84	<0.0001	0.038
Planning target volume coverage	-0.474 \pm 0.169	0.62	0.0050	0.025
Organs at risk and normal tissue sparing	1.003 \pm 0.202	2.73	<0.0001	0.087

^a e represent the odds ratio of the observer preferring the adjusted treatment plan versus the clinical.

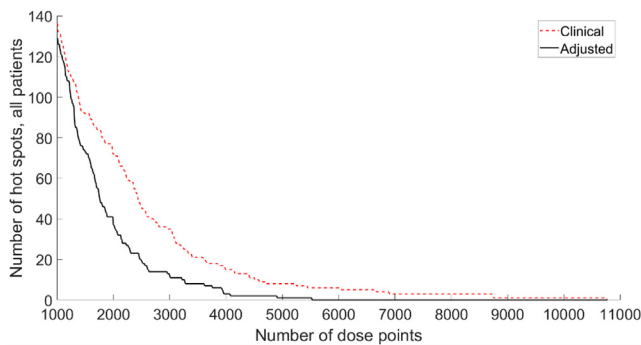


Fig. 2. Histogram over number of hot spots for the whole patient set as function of their sizes as measured in number of dose points (proportional to volume). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

plan and presented as frequencies: Figure 4. For the homogeneity and the normal tissue sparing, there is a shift toward favoring the adjusted treatment plan, and for the PTV there is a shift toward favoring the clinical treatment plan. As seen in Table 3, these trends correspond to sig-

nificant differences in the formal statistical testing with mixed-effects ordinal logistic regression. The McFadden pseudo R² are very low, indicating that there was limited agreement among the observers which is also indicated by the low Fleiss kappa. The preference for the adjusted plan to the clinical plan was also statistically significant. All the observers' answers can be seen in the supplementary material.

The RATING score (22) for the treatment plan comparison is 91%. The RATING score sheet can be found in the supplementary material.

Discussion

As shown in the quantitative analysis of the DVH parameters, the optimization tool preserves the PTV coverage and improves the spatial properties of the treatment plans, resulting in smaller high dose volumes. However, in the observer study the PTV coverage was graded significantly worse. This could be because the locations of the low dose volumes were important to the observers, while the opti-

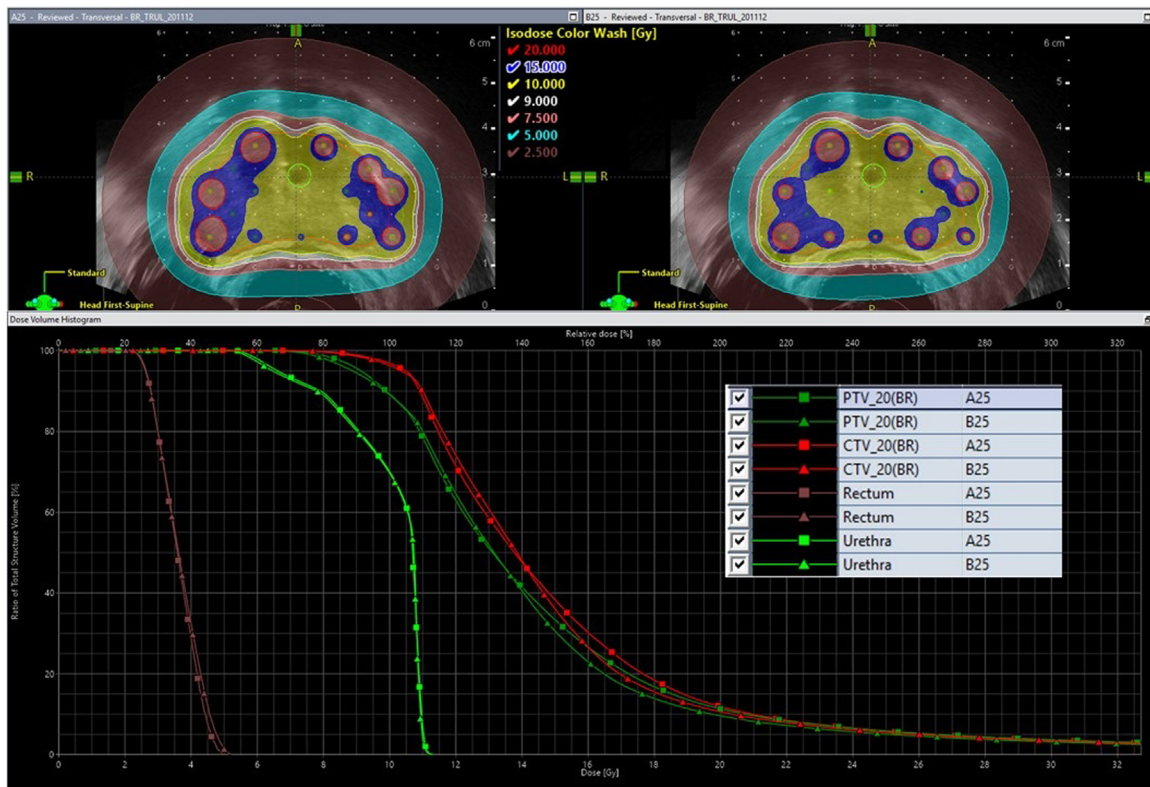


Fig. 3. An example of the treatment plans in the study, patient number 25. The top left image showing one slice of treatment plan A (clinical) and top right image showing the same slice of treatment plan B (adjusted). A dose volume histogram for both treatment plans is shown in the bottom image, with the colors of the structures described. Note the difference of the 20 Gy isodoses (red isodoses) and the difference in the high doses in the dose volume histogram. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

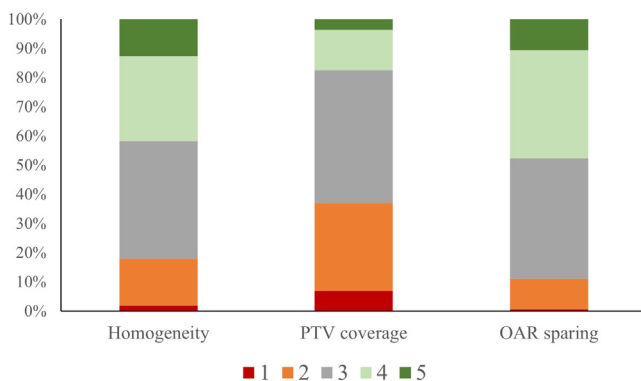


Fig. 4. Frequency diagram for the gradings. The grading scales are: 1 – Confident that clinical treatment plan is better than the adjusted; 2 – Suspect that clinical treatment plan is better than the adjusted; 3 – Treatment plans are equal; 4 – Suspect that clinical treatment plan is worse than the adjusted; and 5 – Confident that clinical treatment plan is worse than the adjusted. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

mization algorithm does not have any location preferences. It can also be seen that there is a significant difference in $V_{100\%}$ for the CTV, with the adjusted treatment plan per-

forming slightly worse. The visual grading was performed on the PTV, but the observer's choices might also be affected by the CTV coverage. The difference in $V_{100\%}$ for CTV is very small – only 0.21% for the mean value – but even minor differences between the treatment plans can affect the grading. It can be seen in Fig. 4 that the most frequently chosen gradings are 2 and 3, indicating that the observer noticed a change for the worse or grades the plans equally with respect to PTV. The prioritization between target coverage and OAR sparing does not seem to be consistent between the observers.

The treatment plans differ by little with respect to $D_{0.1cc}$ and $D_{10\%}$ for urethra and $V_{100\%}$ for CTV. It could be questioned whether these small changes are of clinical importance, even though they are statistically significant. The differences for the hot spots were larger, which shows that it is possible to achieve a more homogenous treatment plan than that obtained by the initial inverse planning. As mentioned before, the clinical impact of hot spots is not proven for prostate cancer. Other anatomical treatment sites such as cervical cancer with more sensitive organs surrounding the target and the tongue being a functional organ, could have larger clinical benefit of the postadjustment tool. In particular, interstitial breast treatment with HDR BT could

benefit from this postadjustment tool, as it has been shown that reducing V_{150} can reduce the incidence rate of fat necrosis (11).

The postadjustment tool was tuned to retain the target coverage and the OAR doses, and to find a more homogeneous treatment plan, with smaller hot spots. The tuning of the optimization tool should be done to match each clinics' needs.

In a treatment plan it is sometimes desirable to give a focal boost, that is a higher dose, to a contoured region such as a dominant intraprostatic lesion. The current post-adjustment tool could then, after focal boost achieved by initial inverse planning, be tuned to focus on minimizing the hot/cold spots in that volume relative to the higher dose levels prescribed there. It is also possible to contour regions that are more sensitive and instruct the initial inverse planning to focus more on such regions (e.g., peripheral zone of prostate), and then use the postadjustment tool to decrease $V_{150\%}$, $V_{200\%}$, and the contiguous hot spots there. The present recommendations by GEC/ESTRO is to use focal boost for prostate only in clinical trials (24). Further, the recommendations in our national guidelines are, as specified above, to give a uniform dose to the target volume, which is what the observers are used to seeing. This study, aimed to evaluate the tool in a clinical environment, is therefore focused on reducing hot spots in this one target volume. Finally, the observer study is easier to carry out if it is as close as possible to the clinical decision making the observers are used to.

The interobserver reliability of this observer study was poor. This could be interpreted as different clinics placing different emphasis on different aspects of a treatment plan, which could explain how the same difference in, for example, homogeneity is graded differently by the observers. In the audit by Dohlmair et al. (31), some clinics have planning aims for V_{150} , but most do not state these types of planning aims. The audit (31) also shows that the number of catheters for the same prostate case differs between 14 and 18, and that $V_{200\%}$ for CTV depends on the number of catheters. The observers could, therefore, be more used to a certain number of hot spots from their clinical treatment plans. The two observers from the clinic where the original treatment plans come from clearly tended to choose the clinical treatment plan, presumably since they resemble the plans that they are used to approving clinically. One of the local observers did indeed in all but three cases choose the clinical treatment plan. The subgroup analysis where the two local observers were removed did not, however, show more agreement in the interobserver reliability analysis. Maree et al. (18) concluded that for three physicians from the same clinic the preferred treatment plan was observer-dependent. The number of observers participating in an observer study is important. Even in this study, where the observer agreement is poor, the rather high number of observers is a strength, as it could have been harder to spot this disagreement with too few observers.

Observer studies can be an important part of a treatment plan comparison, as the properties of an optimal treatment plan cannot be defined in terms of DVH parameters only. Most of the radiobiological models do not take the spatial properties of a treatment plan into account, giving the same result for two treatment plans with the same DVH (10). But different parts of the PTV can have different importance for the observer. Prostate cancer is more often located in the peripheral zone, making the coverage of that volume more important. Maree et al. (18) showed that their observers judged the treatment plans on “volumes with dose higher than 200%, the activation of dwell positions near the OARs, and the location of areas where the target was not covered” in addition to the DVH parameters. Kyrroudi et al. (16) used an observer study to find the optimal part of the Pareto front for EBRT prostate cancer treatments, showing that the observers preferred treatment plans with some PTV under-dosage, since these spared the rectum.

TRAK is a parameter that is often used to compare treatment plans in cervical cancer. This parameter is an estimate of the intensity of the treatment and is independent of contouring uncertainties (32). A lower TRAK can be translated to a smaller total volume being irradiated, which, if the target coverage is the same, yields less radiation to normal tissue. The TRAK for the adjusted treatment plans were significantly lower than those of the clinical plans, which would therefore have given the patients less dose to the normal tissue.

A limitation of the observer study is that it did not include any information about the individual patients, apart from their anatomy. Information about cancer stage and localization of the positive biopsies could have influenced the grading of the treatment plans. Such information could also have influenced the treatment planning in the clinical setting, because of the importance of good coverage where the positive biopsies are located.

Conclusion

The postadjustment optimization tool improves the spatial properties of treatment plans and lowers urethral dose. The three largest hot spots defined as $\geq 200\%$ continuously over-dosed volumes were reduced significantly, and so were $V_{150\%}$ and $V_{200\%}$. The observer study confirmed the conclusions from the analysis of the DVH parameters for all aspects but the PTV coverage. The PTV coverage was, according to the observer study, worse after the postadjustment, showing the importance of not relying solely on DVH parameters when performing a treatment plan comparison.

Acknowledgment

We are grateful to all oncologists who participated in the evaluation of the treatment plans: Ann Henry, Bengt Johansson, Enrique Castellanos, Gabriel Lindahl,

Jan Rzepecki, Lennart Åström, Måns Agrup and Oscar Derke. We are also grateful to Lars Valter (Forum Östergötland, Linköping University) for providing statistical advice regarding *t* tests and kappa analysis.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.brachy.2022.12.008](https://doi.org/10.1016/j.brachy.2022.12.008).

References

- [1] Palmer A, Bradley D, Nisbet A. Physics-aspects of dose accuracy in high dose rate (HDR) brachytherapy: source dosimetry, treatment planning, equipment performance and in vivo verification techniques. *J Contemp Brachyther* 2012;4:81–91.
- [2] Dinkla AM, van der Laarse R, Kaljouw E, et al. A comparison of inverse optimization algorithms for HDR/PDR prostate brachytherapy treatment planning. *Brachytherapy* 2015;14:279–288.
- [3] Morén B, Larsson T, Tedgren ÅC. Optimization in treatment planning of high dose-rate brachytherapy — review and analysis of mathematical models. *Med Phys* 2021;48:2057–2082.
- [4] Panettieri V, Smith RL, Mason NJ, Millar JL. Comparison of IPSA and HIPO inverse planning optimization algorithms for prostate HDR brachytherapy. *J Appl Clin Med Phys* 2014;15:256–266.
- [5] Holm Å, Larsson T, Carlsson Tedgren Å. Impact of using linear optimization models in dose planning for HDR brachytherapy. *Med Phys* 2012;39:1021–1028.
- [6] Vikram B, Deore S, Beitler JJ, et al. The relationship between dose heterogeneity ('hot' spots) and complications following high-dose rate brachytherapy. *Int J Radiat Oncol Biol Phys* 1999;43:983–987.
- [7] Duchesne GM, Das R, Teye W, See A. Dose distribution and morbidity after high dose rate brachytherapy for prostate cancer: influence of V150 and V200 parameters. *Australas Radiol* 2002;46:384–389.
- [8] Chichef A, Kanikowski M, Skowronek J. Vital role of volume and number of needles in HDR brachytherapy (HDR-BT) of prostate cancer. *J Contemp Brachyther* 2009;1:145–150.
- [9] Aydogan B. A dosimetric analysis of intensity-modulated radiation therapy (IMRT) as an alternative to adjuvant high-dose-rate (HDR) brachytherapy in early endometrial cancer patients. *Int J Radiat Oncol Biol Phys* 2006;65:266–273.
- [10] Mavroidis P, Katsilieri Z, Kefala V, et al. Radiobiological evaluation of the influence of dwell time modulation restriction in HIPO optimized HDR prostate brachytherapy implants. *J Contemp Brachyther* 2010;2:117–128.
- [11] Garsa AA, Ferraro DJ, Dewees T, et al. Analysis of fat necrosis after adjuvant high-dose-rate interstitial brachytherapy for early stage breast cancer. *Brachytherapy* 2013;12:99–106.
- [12] Strnad V, Major T, Polgar C, et al. ESTRO-ACROP guideline: interstitial multi-catheter breast brachytherapy as Accelerated Partial Breast Irradiation alone or as boost - GEC-ESTRO Breast Cancer Working Group practical recommendations. *Radiother Oncol* 2018;128:411–420.
- [13] Poder J, Whitaker M. Robustness of IPSA optimized high-dose-rate prostate brachytherapy treatment plans to catheter displacements. *J Contemp Brachyther* 2016;8:201–207.
- [14] Morén B, Larsson T, Tedgren ÅC. A mathematical optimization model for spatial adjustments of dose distributions in high dose-rate brachytherapy. *Phys Med & Biol* 2019;64:225012.
- [15] Petersson K, Engellau J, Nilsson P, et al. Treatment plan comparison using grading analysis based on clinical judgment. *Acta Oncologica* 2013;52:645–651.
- [16] Kyroudi A, Petersson K, Ozsahin E, et al. Exploration of clinical preferences in treatment planning of radiotherapy for prostate cancer using Pareto fronts and clinical grading analysis. *Phys Imaging Radiat Oncol* 2020;14:82–86.
- [17] Oud M, Kolkman-Deurloo IK, Mens JW, et al. Fast and fully-automated multi-criterial treatment planning for adaptive HDR brachytherapy for locally advanced cervical cancer. *Radiother Oncol* 2020;148:143–150.
- [18] Maree S, Luong H, Kooreman E, et al. Evaluation of bi-objective treatment planning for high-dose-rate prostate brachytherapy—A retrospective observer study. *Brachytherapy* 2019;18:396–403.
- [19] Bélanger C, Poulin É, Aubin S, et al. Inter-observer evaluation of a GPU-based multicriteria optimization algorithm combined with plan navigation tools for HDR brachytherapy. *Brachytherapy* 2022;21:551–560.
- [20] European Communities. Report EUR 16262 EN (1996) European guidelines On Quality Criteria For Computed Tomography. Office for Official Publications of the European Communities. 1996.
- [21] Smedby Ö, Fredrikson M. Visual grading regression: analysing data from visual grading experiments with regression models. *Br J Radiol* 2010;83:767–775.
- [22] Hansen CR, Crijns W, Hussein M, et al. Radiotherapy Treatment planning study Guidelines (RATING): a framework for setting up and reporting on scientific treatment planning studies. *Radiother Oncol* 2020;153:67–78.
- [23] Regionalt cancercentrum. National care program prostate cancer v5.0. 2020.
- [24] Henry A, Pieters BR, F André Siebert, Hoskin P. GEC-ESTRO ACROP prostate brachytherapy guidelines. *Radiother Oncol* 2022;167:244–251.
- [25] Baltas D, Kolotas C, Germani K, et al. A conformal index (COIN) to evaluate implant quality and dose specification in brachytherapy. *Int J Radiation Oncology Biol Phys* 1998;40:515–524.
- [26] Saw CB, Suntharalingam N. Quantitative assessment of interstitial implants. *Int J Radiat Oncol Biol Phys* 1991;20:135–139.
- [27] Manikandan A, Sarkar B, Rajendran VT, et al. Role of step size and max dwell time in anatomy based inverse optimization for prostate implants. *J Med Phys* 2013;38:148–154.
- [28] Morén B, Bokrantz R, Dohlmair F, et al. Technical note: Evaluation of a spatial optimization model for prostate high dose-rate brachytherapy in a clinical treatment planning system. *Med Phys* 2023;1–6.
- [29] Smedby Ö, Fredrikson M, De Geer J, Sandborg M., Visual grading regression with random effects, Proc. SPIE 8318, Medical Imaging 2012: Image Perception, Observer Performance, and Technology Assessment, 831805 (22 February 2012).
- [30] Hemmert GAJ, Schons LM, Wieseke J, Schimmelpfennig H. Log-likelihood-based pseudo-R² in logistic regression: deriving sample-sensitive benchmarks. *Sociolog Methods & Res* 2018;47:507–531.
- [31] Dohlmair F, Johansson S, Larsson T, et al. An audit of high dose-rate prostate brachytherapy treatment planning at six Swedish clinics. *J Contemp Brachyther* 2021;13:59–71.
- [32] Nkiwane KS, Andersen E, Champoudry J, et al. Total reference air kerma can accurately predict isodose surface volumes in cervix cancer brachytherapy. A multicenter study. *Brachytherapy* 2017;16:1184–1191.