Radiotherapy and Oncology 133 (2019) 198–204

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Independent knowledge-based treatment planning QA to audit Pinnacle autoplanning

Tomas M. Janssen *, Martijn Kusters, Yibing Wang, Geert Wortel, Rene Monshouwer, Eugène Damen, Steven F. Petit

Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam; Department of Radiation Oncology, Radboud University, Nijmegen; and Department of Radiation Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands

ARTICLE INFO

Article history:
Received 4 May 2018
Received in revised form 26 October 2018
Accepted 28 October 2018
Available online 15 November 2018

Keywords:
Knowledge based planning
Autoplanning

ABSTRACT

Background and purpose: With the advent of automatic treatment planning options like Pinnacle's Autoplanning (PAP), the challenge arises how to assess the quality of a plan that no dosimetrist did work on. The aim of this study was to assess plan quality consistency of PAP prostate cancer patients in clinical practice.

Materials and methods: 100 prostate cancer patients were included from NKI and 129 from RadboudUMC (RUMC). Per institute a previously developed [1] treatment planning QA model, based on overlap volume histograms, was trained on PAP plans to predict achievable dose metrics which were then compared to the clinical PAP plans. A threshold of 3 Gy (DVH dose parameters)/3% (DVH volume parameters) was used to detect outliers. For the outlier plans, the PAP technique was adjusted with the aim of meeting the threshold.

Results: The average difference between the prediction and the clinically achieved value was <0.5 Gy (mean dose parameters) and <1.2% (volume parameters), with standard deviation of 1.9 Gy/1.5% respectively. We found 8% (NKI)/25% (RUMC) of patients to exceed the 3 Gy/3% threshold, with deviations up to 6.7 Gy (mean dose rectum) and 6% (rectal wall V64Gy). In all cases the plans could be improved to fall within the thresholds, without compromising the other dose metrics.

Conclusion: Independent treatment planning QA was used successfully to assess the quality of clinical PAP in a multi-institutional setting. Respectively 8% and 25% suboptimal clinical PAP plans were detected that all could be improved with replanning. Therefore we recommend the use of independent treatment plan QA in combination with PAP for prostate cancer patients.

Conventional inverse treatment planning is a manual trial and error process [2]. This makes the treatment planning process in practice time consuming and subject to a substantial amount of variation between dosimetrists [3,4]. To increase treatment planning efficiency and plan consistency several different 'auto-planning' approaches have emerged in the past years [5–9]. The basic aim of all these approaches is that a single auto-planning protocol results in good plans for all patients for a given treatment site, without further adjustments; in contrast to conventional planning where a treatment site class solution needs to be manually adjusted on a per patient basis.

Pinnacle autoplanning (PAP) [7] is one of the commercially available solutions for auto-planning using a template-based optimization tool that mimics the iterative optimization steps a skilled dosimetrist would undertake during manual planning. PAP has been shown to yield good quality plans with respect to manual planning for several treatment sites [10–13]. However in daily practice PAP is a 'black box' solution that produces a plan and for a dosimetrist it is very hard, if not impossible, to judge whether the resulting plan could potentially be further improved.

Indeed, since it is hard to objectively judge an autoplan, the question arises, if PAP, or any other available autoplanning solution, consistently generates high quality treatment plans for all patients. If one in several patients receives a suboptimal plan, this may be unnoticed, because of the issue mentioned above. Assessing the consistency of PAP is therefore crucial for a widespread clinical deployment, but to our knowledge the use of PAP in clinical practice has never been properly addressed for a large number of patients.

Therefore the aim of this study was to assess and quantify the performance of PAP in daily clinical practice for a large group of prostate cancer patients. Performance was expressed by the

* Corresponding author at: The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.
E-mail address: t.janssen@nki.nl (T.M. Janssen).

https://doi.org/10.1016/j.radonc.2018.10.035
0167-8140/

number of PAP plans that were detected as suboptimal by an independent tool for treatment planning QA and that indeed could be improved with replanning.

Materials and methods

A previously developed independent plan QA tool, based upon overlap volume histograms (OVH) [1], was used to detect PAP plans that were outliers in terms of achieved DVH parameters. The outliers' plans were then replanned to confirm that indeed the PAP plans could have been improved. In order to make the results more generally applicable, we performed the analysis independently at two different hospitals in the Netherlands (Netherlands Cancer Institute, Amsterdam: NKI; Radboud University Medical Center, Nijmegen: RUMC), using two different treatment protocols, with a total of 229 patients.

Patient data from NKI

The data for this study from the NKI consisted of 100 clinical PAP treatment plans from patients treated consecutively in 2016–2017 for prostate cancer CTVs consisted of the entire prostate (N = 29) or the prostate including the seminal vesicles (N = 71). CTVs were delineated on the planning CT (3 mm slice thickness), matched to T2 weighted MRI. Dose was prescribed using a simultaneous boost technique of 35 × 2.2 Gy to the CTV + 4 mm and 35 × 2.0 Gy to the CTV + 7 mm. Organs at risk were the delineated rectum (from anus to ascending colon), rectal wall (2 mm thickness), anal sphincter (the most caudal 3 cm of the rectum) and femoral heads. The bladder was, according to local protocol, not taken into account as OAR.

PAP treatment planning at NKI

All plans were dual arc VMAT plans of a partial rotation over 280 degrees, excluding the dorsal part of the arc, using 10MV photons. All treatment plans were created using the same PAP protocol (Pinnacle version 9.10, Philips Health Care, Best, The Netherlands), followed by an in-house Pinnacle script for final adjustments. This script adds several cost functions to the final PAP and runs one more round of optimization. The aim of this script is to ensure target coverage, improve conformity, remove hotspots near the rectum, and meet the DVH constraint for the femoral heads. The specific details of this script are provided in the Supplementary table.

All plans were manually checked by a dosimetrist and physician before clinical delivery and if judged necessary, manual adjustments could have been made. In clinical practice this happens in around 10% of cases, only to suppress hotspots in the PTV or to improve lateral conformality.

A summary of the autoplanning technique is given in Table 1a. Plan evaluation parameters are summarized in Table 2a. For this study we examined the rectal wall V64Gy, the mean dose on the anal sphincter and the 'soft constraint' on the mean dose on the rectum. The rectal wall V75Gy was not taken into account, because its value is primarily determined by the overlap between the rectum and the PTV and thus not prone to large variations in treatment plan quality (this was verified explicitly; data not shown).

Table 1a
Pinnacle autoplanning technique at the NKI.

ROI	Type	Goal	Priority
PTV Low dose – PTV high dose	Target	7000 cGy	NA
PTV high dose	Target	7700 cGy	NA
Rectal wall	Mean dose	0 cGy	High
PTVring	Max Dose	6650 cGy	High
Anal sphincter	Mean dose	0 cGy	High
Rectal wall	Max DVH	7500 cGy ; 7%.	High
Rectum – (PTV + 5 mm)	Max Dose	0 cGy	Medium

Table 1b
Pinnacle autoplanning technique at the RUMC.

ROI	Type	Goal	Priority
PTV	Target	7500 cGy	NA
Rectal wall	Max DVH	6000 cGy; 30%	High
Rectal wall	Max DVH	3000 cGy; 80%	Low
Rectal wall	Mean dose	4500 cGY	High
Bladder	Max DVH	6000 cGy < 30%	High
Anal wall	Mean dose	3000 cGy	High
Rectal + Anal wall	Max DVH	6000 cGy; 40%	High
Rectal + Anal wall	Max DVH	3000 cGy; 80%	High
ex-ROI	Max DVH	4500 cGy; 15%	Low

Table 2a
Plan evaluation criteria at the NKI.

ROI	Metric	Criterion
PTVs	V95%	>99%
	V107%	<1%
Rectal wall	V75Gy	<10%
	V64Gy	<35%
Anal sphincter	Dmean	<45 Gy
Femur Head	Dmax	<50 Gy
Soft constraints		
Rectum	Dmean	'ALARA'
External	conformality 90% isodose	'Conformal'

Patient data from RUMC

The data for this study from RUMC consisted of 129 clinical PAP treatment plans from 2015 to 2017, treated according to local protocol which is based on national guidelines [14]. CTVs were delineated on the planning CT (3 mm slice thickness) and consisted of the entire prostate (N = 15), prostate plus part of the seminal vesicles (N = 63) or prostate with complete seminal vesicles (N = 51). CTV to PTV margin is 5 mm for dorsal direction and 7 mm for other directions. A dose of 28 × 2.5 Gy was prescribed to the PTV. All patients were treated using an endorectal balloon (ERB), filled with 100 ml of air, in order to spare the posterior rectal and anal wall and reduce toxicity for these organs [15]. Organs at risk were the anal wall (Awall), rectal wall (Rwall) and the bladder.

PAP treatment planning RUMC

All plans were single arc VMAT plans using a partial rotation over 190 degrees, excluding the dorsal part of the arc using 10 MV photons. All treatment plans were created using the same PAP protocol (Pinnacle version 9.10, Philips Health Care, Best, The Netherlands). All plans were manually checked by a dosimetrist and physician before clinical delivery but not manually tweaked afterwards.

A summary of the autoplanning technique is given in Table 1b. Plan evaluation parameters are summarized in Table 2b. For this study we examined the rectal wall V64Gy, the mean dose on the rectal wall and the mean dose on the anal wall.

DVH prediction

Per institute, patients were randomly divided 50–50 in a training and validation group. The training patients were used to train a recently published KB treatment planning QA model [1]. The KB model uses the principal components of the Overlap Volume Histogram (OVH), as the OAR's geometrical features to predict the principal components of the achievable DVH, using a support

Table 2b
Plan evaluation criteria at the RUMC.

ROI	Metric	Criterion
PTVs	V95%	>99%
	V105%	<5%
Rectal Wall	V60Gy	<30%
	V30Gy	<80%
	Dmean	<45 Gy
Bladder	V60Gy	<30%
Anal wall	Dmean	<30 Gy
Femur Head	V50Gy	<10%
Rectal + Anal wall	V60Gy	<40%
Rectal + Anal wall	V30Gy	<80%

vector regression model. The KB model was used to predict the achievable DVH metrics for the validation patients. Because of the differences in the planning protocol and the use of an endorectal balloon in the RUMC, the model was trained independently for both institutes. We focused on the DVH parameters of the rectum (rectum, anal wall, anal sphincter and rectal wall), being the most critical structure for plan evaluation. The bladder was not considered an OAR in clinical practise at NKI and the femoral head constraints were always met. Therefore the bladder and femoral heads were not added to the analysis. Plans were defined as outliers when the achieved dose was more than 3 Gy (mean rectum and anus dose) or 3% (dose volume parameters) higher than the dose predicted by the KB QA model. The 3 Gy and 3% thresholds correspond roughly to the accuracy of the KB model (2 SD) as determined previously [16].

Using the model, plans were identified that could potentially be improved. For these patients, the PAP technique would be manually adjusted by changing the OAR goals, to aim for a plan that would fall within the expected predicted threshold.

Results

Results NKI

Fig. 1 shows the results of the DVH prediction in the evaluation cohort for the different metrics. The average mismatch (predicted – achieved) ± standard deviation for the evaluated parameters was rectum Dmean: −0.1 Gy ± 1.9 Gy; anal sphincter Dmean: +0.2 Gy ± 1.3 Gy and rectal wall V64Gy −0.7% ± 1.4%.

Of the 50 validation patients, plans of 5 patients (7 DVH metrics) exceeded the action level of 3 Gy/3%. Of these 5 patients one had a double sided hip prosthesis and an alternative, non-AP, planning strategy was used clinically. No other such cases were present in the dataset and this case was excluded from further analyses. The metrics exceeded were the rectum Dmean (3x) and rectal wall V64Gy (1x), with deviations up to 6.7 Gy (mean dose rectum) and 3.1% (rectal wall V64Gy) (Table 3a).

The remaining 4 cases (8%) could be improved such that all dose metrics became within the 3 Gy/3% interval, without compromising the other metrics such that they would have exceeded the threshold (see Table 3a). The dose distributions of an example case are shown in Fig. 3.

Results RUMC

Fig. 2 shows the results of the DVH prediction in the evaluation cohort for the different metrics. The average mismatch (predicted – achieved) ± standard deviation for the evaluated parameters was rectal wall Dmean: −0.1 Gy ± 2 Gy; anal wall Dmean: −0.5 Gy ± 2.1 Gy; and rectal wall V64Gy −1.2% ± 1.6%.

Of the 65 validation patients, plans of 16 (25%) patients (20 DVH metrics) exceeded the action level of 3 Gy/3%. The metrics exceeded were the rectum Dmean (4x), rectal wall V64Gy (8x) and anal wall Dmean (8x), with deviations up to 6.3 Gy (mean dose anal wall) and 6.0% (rectal wall V64Gy). All cases could be improved such that all dose metrics became within the 3 Gy/3% interval, without compromising the other metrics such that they would have exceeded the threshold (see Table 3b).

Discussion

In this study we assessed and quantified the performance of PAP using an independent KB treatment planning QA model. In total 229 (split equally in training and evaluation sets) clinical prostate treatment plans generated with PAP at two different institutes were analysed. We found that, depending on the institute, 8% to 25% of PAP plans were identified by the planning QA model as suboptimal an all these plans could indeed be approved.

Plan quality of plans generated using PAP has been assessed in the literature. For example Hansen et al [10] found for 29 out of 30 head and neck patients that PAP, with a 'minor manual fine-tuning' after autoplanning, resulted in superior plans compared to manual planning, expressed by a more homogeneous target dose and a decrease in OAR doses of up to 6.5 Gy. Similar results have been obtained for oesophageal cancer [17]. Using PAP for Prostate cancer was studied in [18], where it was found for 23 patients that PAP was equivalent or better compared to manual planning. Moreover it was shown explicitly that PAP reduces inter-observer dependency. However for a safe, large scale, clinical deployment of PAP, the stability of the algorithm also needs to be assessed over a large range of patients from daily clinical practice. The need for such a stability analysis arises from the observation that PAP (like other autoplanning solutions) basically is a black box that produces a plan, without any explicit quality metrics. Of course a dosimetrist or physician can then check whether the plan is clinically acceptable (e.g. whether all constraints are met), but checking if this is indeed the *best* plan for this patient turns out to be very hard [3]. In practice one way to assess whether or not a plan can be further improved is by manual trial and error, however this was precisely the time-consuming and user dependent process autoplanning tried to avoid.

Knowledge based planning [8,9,19–21] has been suggested as a tool both for treatment plan quality QA and as input for autoplanning (commercialized by Varian as Rapidplan [22]). These approaches all work on the same premise: based on geometric characteristics of the patient the dose distribution of a new patient can be predicted, by comparing the geometry with a database of previously treated patients. Using a previously developed prediction model based on overlap volume histograms [1], we found an average deviation between the prediction and the clinically achieved value of the PAP plans of <0.5 Gy (mean dose parameters) and <1.2% (volume parameters). The standard deviation of the difference between the predictions was, averaged over the mean dose parameters, 1.6 Gy (NKI)/2.1 Gy (RUMC) and, averaged over the volume parameters, 1.4% (NKI)/1.6% (RUMC). We can compare this with the equivalent numbers from [16] (1.3 Gy and 1.4% respectively), where the same KB model was used on a set of uniformly generated, Pareto optimal plans. These results can therefore be interpreted as the upper bound of the predictive power of the model. Since we found that the results of the PAP are in a similar range, we conclude that PAP produces very stable results.

There are several key differences between the NKI and RUMC PAP. First of all, as can be seen in Table 1, the input for PAP differs between the institutions. The NKI approach here is to ask for as little dose as possible in the OAR (e.g. rectum Dmean: 0 cGy), where RUMC uses more realistic criteria (e.g. rectum Dmean: 4500 cGy). Asking for an (unrealistic) 0 cGy in the OAR might force PAP to

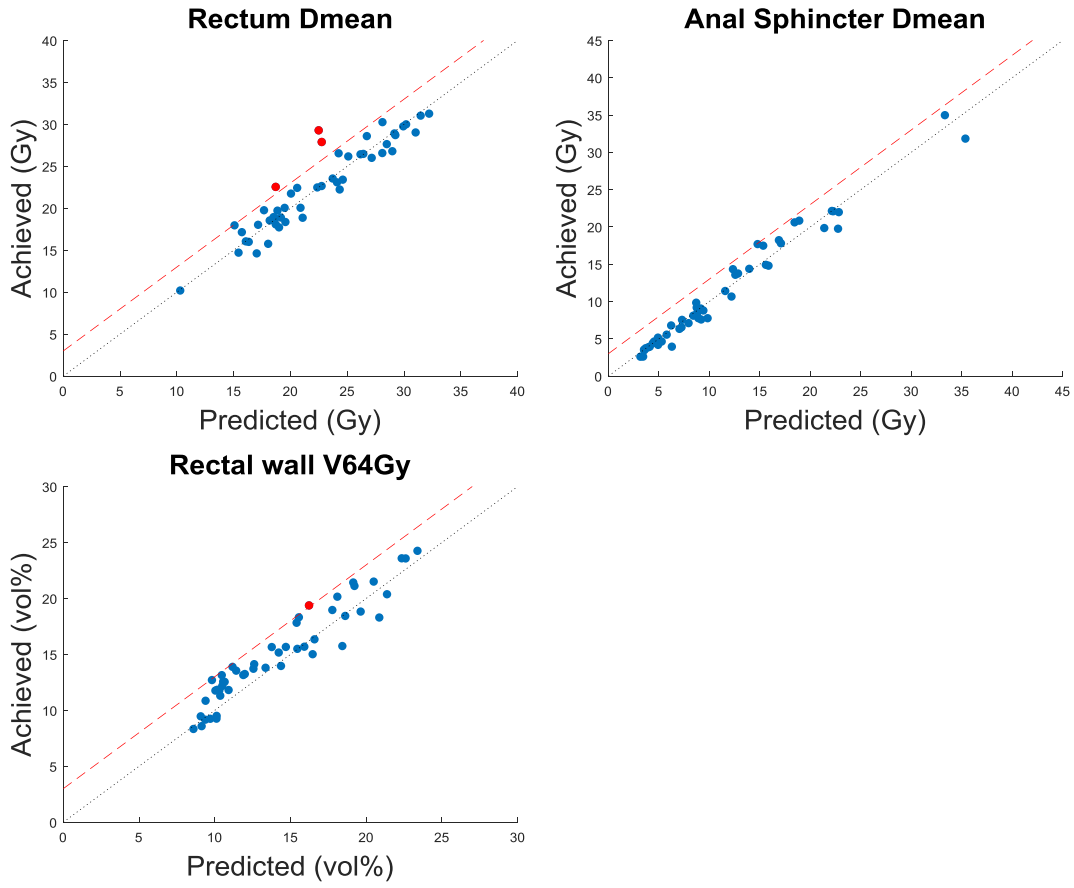

Fig. 1. Achieved versus predicted dose metrics for the NKI dataset. The red dashed line indicates the 3 Gy/3vol% threshold. The black dotted line indicates unity. Outliers are indicated in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

always push similar on the OAR, where if an explicit target goal of 4500 cGy is asked, PAP won't push as hard as soon as this dose level is reached, leading to a different trade-off.

Another difference in the PAP technique is the number of auto-planning goals. Apart from the targets, NKI uses 5 where RUMC uses 8. This is however in accordance with the somewhat longer list of explicit evaluation criteria. It is not unreasonable to argue that these differences could lead to a more stable solution in the NKI approach: less objectives make a simpler and more robust optimization problem. How this works out for PAP, where over six iterations multiple optimization objectives were added for each auto-planning goal is at the moment not clear. A second key difference is that RUMC uses just PAP, with no adjustments after the plan is ready, where NKI first employs a Pinnacle script to fine-tune the plan and then allows for a manual adjustment by the dosimetrist. One could argue that either this script, or some manual fine-tuning could remove some of the outliers and make the plans more homogeneous, though this is probably not the case for the manual fine tuning. First of all, inter-dosimetrist variation is known to occur and in fact is often hauled as one of the arguments for autoplanning and second of all, one of the outliers found in the NKI dataset was actually a manually tweaked plan. It should also be noted that manually tweaking only occurred in about 10% of cases and focusses mostly on fine tuning target homogeneity. We do not expect this to have a large influence on our conclusions. The script to fine-tune the plan however could very well be a cause of a more homogeneous, predictable dataset. It is worth mentioning that [10] also required some fine-tuning after PAP to obtain the best results in head and neck cancer patients.

From the results and Figs. 1 and 2 it is clear that there are more outlier plans in the RUMC set compared to the NKI set. Moreover we see that most of the outliers in the RUMC set are in the higher dose region. Potentially these outliers could have been avoided by reoptimzing the PAP protocol (i.e. for all patients). Indeed based on these findings the clinical RUMC PAP model has been updated since. This finding demonstrates that (i) finding the best PAP protocol is not straightforward and (ii) that plan QA is not only useful for PAP on an individual patient basis, but could also provide valuable insight for evaluating/reoptimizing PAP protocols.

In all cases, the autoplanning technique could be adjusted such that plans did meet the expected values, without compromising the other metrics. Of course, as in all treatment planning studies, everything comes at a price, so some changes in the other evaluation parameters was seen, but these changes were such that the metrics would still be within the predicted thresholds. Nonetheless it is important to realize that outliers found using this method (that predicts all DVHs independently), does not necessarily mean that the plan was necessarily 'bad', but it does imply that a trade-off between objectives was made that was different from what was expected on the training data.

The fact that all plans that were outside of the pre-set action level could be improved to meet the expected values, implies that in these cases, the single PAP technique used in clinical practice did not reach the expected trade-off, compared to what you would expect based on similar patients. An outlier plan is however not by definition a poor plan: the differences found might not always be clinically relevant and a different trade-off might not be a worse trade-off. However, compared to clinical practice these plans are outliers, and based on the premise that the typical trade-off is

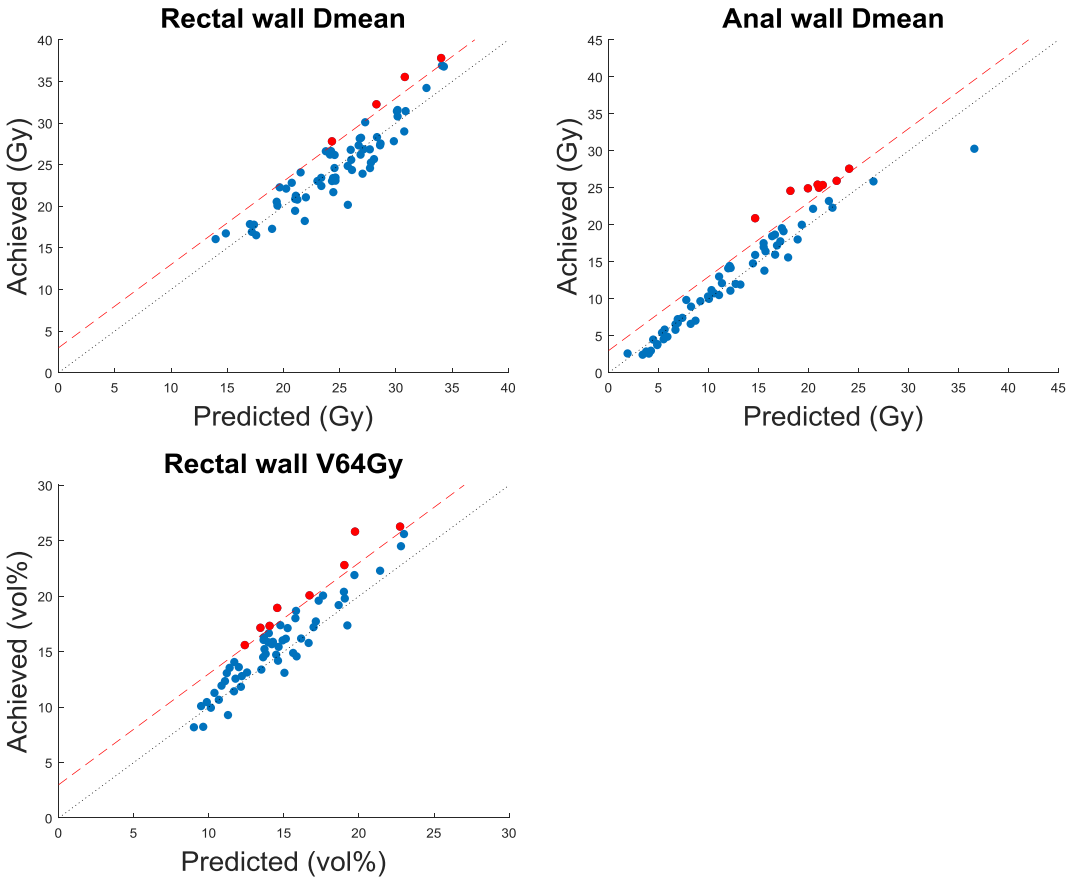

Fig. 2. Achieved versus predicted dose metrics for the RUMC dataset. The red dashed line indicates the 3 Gy/3vol% threshold. The black dotted line indicates unity. Outliers are indicated in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3a
Outliers NKI. All dose metrics that were detected as outliers are presented.

Patient	Metric	Prediction	Achieved clinically	Achieved with replan	Prediction – clinical	Prediction – replan
1	Rectal wall V64Gy [%]	16.3	19.3	18.1	−3	−1.8
2	Rectum Dmean [Gy]	22.5	29.2	21.8	−6.7	0.7
3	Rectum Dmean [Gy]	22.8	27.8	24.4	−5	−1.6
4	Rectum Dmean [Gy]	18.8	22.5	20.5	−3.7	−1.7

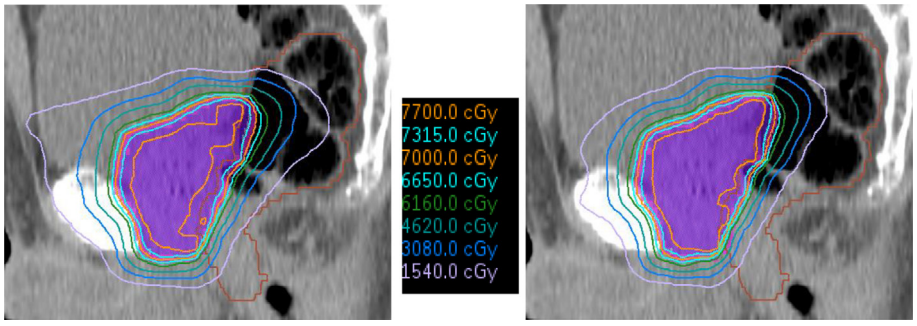

Fig. 3. Clinical plan (left) versus replan after the plan was flagged by the QA tool (right), for one of the NKI patients. The plan was flagged based on the mean rectum dose (clinical: 27.8 Gy; prediction 22.8 Gy; after replan: 24.4 Gy). The high dose PTV is shown in purple, low dose PTV in pink, and rectum in brown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

expected to be the optimal trade-off, the outliers are sub-optimal. Moreover, since these plans were all checked manually in clinical practice, this shows that this sub-optimality was not detected by the dosimetrist or physician. Therefore, the need for an independent QA tool in clinical practice is evident in order to ensure that the best possible plan is generated for each patient. While this work focusses on prostate planning, it should be noted that both the PAP technique, as the treatment protocol (with or without

Table 3b
Outliers RUMC. All dose metrics that were detected as outliers are presented.

Patient	Metric	Prediction	Achieved clinically	Achieved with replan	Prediction – clinical	Prediction – replan
1	Rectal wall Dmean [Gy]	30.8	34.9	27.8	−4.1	3
1	Rectal wall V64Gy [%]	19.8	25	22.4	−5.2	−2.6
2	Rectal wall Dmean [Gy]	28.3	31.5	27.5	−3.2	0.8
2	Rectal wall V64Gy [%]	16.7	19.1	17.2	−2.4	−0.5
3	Rectal wall Dmean [Gy]	24.4	27	25.4	−2.6	−1
3	Rectal wall V64Gy [%]	14.6	17.8	17.2	−3.2	−2.6
4	Rectal wall Dmean [Gy]	34.1	36.9	31	−2.8	3.1
4	Rectal wall V64Gy [%]	22.8	25.2	22.8	−2.4	0
5	Anal wall Dmean [Gy]	14.7	20.6	15.9	−5.9	−1.2
6	Anal wall Dmean [Gy]	22.9	25.6	22.5	−2.7	0.4
7	Anal wall Dmean [Gy]	21.1	24.4	20.7	−3.3	0.4
8	Anal wall Dmean [Gy]	20	24.9	22.8	−4.9	−2.8
9	Anal wall Dmean [Gy]	24.1	27.1	23.2	−3	0.9
10	Anal wall Dmean [Gy]	20.9	25.4	21.5	−4.5	−0.6
11	Anal wall Dmean [Gy]	21.5	25.3	21.9	−3.8	−0.4
12	Anal wall Dmean [Gy]	18.2	24.3	21.1	−6.1	−2.9
13	Rectal wall V64Gy [%]	19.1	21.7	21	−2.6	−1.9
14	Rectal wall V64Gy [%]	12.4	14.6	14.2	−2.2	−1.8
15	Rectal wall V64Gy [%]	13.5	15.9	14.2	−2.4	−0.7
16	Rectal wall V64Gy [%]	14.1	17.2	16.4	−3.1	−2.3

endorectal balloon, one or two dose levels, one or two VMAT arcs, ...) is considerably different between the two institutions and therefore we expect the results in this paper to be representative for PAP in general. Also, since the literature on PAP is quite consistent over different treatment sites [10,17,18], we do not see any reason why PAP for more complex treatment sites would result in a more stable solution. As a matter of fact, due to a larger anatomical variation in tumour location in sites like lung and head and neck, the opposite might be more likely.

In clinical practice protocols and techniques might change over time. In theory the QA model as we described should then be adjusted accordingly, to deal with different trade-offs or dose gradients. To ensure the independence of the QA model however a continuously adapting model is not recommended. The reason being that potentially induced errors due to a new protocol will not be found if the same errors are present in the model.

Since we showed that by adjusting the planning protocol, the predicted values could be met, it is tempting to think that the predicted dose metrics could be directly used to seed the Pinnacle autoplanning technique. In this way one could expect that a possible improvement of all plans is possible, including the non-outlier plans. The current work cannot be used to fully answer that question, since we only replanned plans that were detected as outliers and not non-outlier plans. In this way we did study the specificity of the QA tool, but not the sensitivity. This approach, of seeding PAP with KB predictions, is currently under further investigation.

In this work we showed that planning QA can be used to detect flaws in the automated treatment planning process. Since automation may reduce user awareness which may increase the likelihood of human mistakes, we believe that independent treatment planning QA can be useful for any other autoplanning solution, including Rapidplan that uses a KPB based approach to set its initial objectives [22,23,24] or Erasmus-iCycle [5,15] that produces mathematically Pareto optimal fluence plans as automated input for Monaco.

Finally, in the larger context of workflow and treatment planning automation, one can envision a workflow where treatment planning is performed automatically, but also the independent QA on the plan is performed automatically and a human, manual check is only required for outlier cases. This could increase throughput, and reduce workload on the planning department. Again this highlights the need for an independent QA of autoplanning. This work is a first step in that direction.

Conflict of interest statement

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2018.10.035.

References

[1] Wang Y, Heijmen BJM, Petit SF. Prospective clinical validation of independent DVH prediction for plan QA in automatic treatment planning for prostate cancer patients. Radiother Oncol 2017;125:500–6.
[2] Hunt MA, Hsiung C-Y, Spirou SV, Chui C-S, Amols HI, Ling CC. Evaluation of concave dose distributions created using an inverse planning system. Int J Radiat Oncol Biol Phys 2002;54(3):953–62.
[3] Berry SL, Boczkowski A, Ma R, Mechalakos J, Hunt M. Interobserver variability in radiation therapy plan output: results of a single-institution study. Pract Radiat Oncol 2016;6:442–9.
[4] Wright K et al. Improving plan quality for prostate volumetric-modulated arc therapy. Med Dosim 2017;42:348–56.
[5] Voet PWJ, Dirkx MLP, Breedveld S, Al-Mamgani A, Incrocci L, Heijmen BJM. Fully automated volumetric modulated arc therapy plan generation for prostate cancer patients. Int J Radiat Oncol Biol Phys 2014;88:1175–9.
[6] Purdie TG, Dinniwell RE, Letourneau D, Hill C, Sharpe MB. Automated planning of tangential breast intensity-modulated radiotherapy using heuristic optimization. Int J Radiat Oncol Biol Phys 2011;416:1–9.
[7] Hazell I et al. Automatic planning of head and neck treatment plans. J Appl Clin Med Phys 2016;17:5901.
[8] Petit SF et al. Increased organ sparing using shape-based treatment plan optimization for intensity modulated radiation therapy of pancreatic adenocarcinoma. Radiother Oncol 2012;102:38–44.
[9] Appenzoller LM, Michalski JM, Thorstad WL, Mutic S, Moore KL. Predicting dose-volume histograms for organs-at-risk in IMRT planning. Med Phys 2012;39:7446–61.
[10] Hansen CR et al. Automatic treatment planning improves the clinical quality of head and neck cancer treatment plans. Clin Transl Radiat Oncol 2016;1:2–8.
[11] Dawn G et al. Initial evaluation of automated treatment planning software. J Appl Clin Med Phys 2016;17:331–46.
[12] Wu B et al. Cross-institutional knowledge-based planning (KBP) implementation and its performance comparison to Auto-Planning Engine (APE). Radiother Oncol 2017;123:57–62.
[13] Kusters M, Bzdusek K, Kumar P, Van Kollenburg P, Kunze-busch M, Kaanders H. Automate IMRT Planning in pinnacle: a study in head-and-neck cancer. Strahlenther Onkol 2017;193:1077–8.
[14] T. de Reijke et al., "https://oncoline.nl/prostaatcarcinoom," 2016.
[15] Smeenk RJ, Van Lin ENJT, Van Kollenburg P, McColl GM, Kunze-Busch M, Kaanders JHAM. Endorectal balloon reduces anorectal doses in post-prostatectomy intensity-modulated radiotherapy. Radiother Oncol 2011;101:465–70.
[16] Wang Y, Breedveld S, Heijmen B. Evaluation of plan quality assurance models for prostate cancer patients based on fully automatically generated Pareto-optimal treatment plans. Phys Med Biol 2016;61:4268.

[17] Hansen CR et al. Automatic treatment planning facilitates fast generation of high-quality treatment plans for esophageal cancer. Acta Oncol. (Madr) 2017;56:1495–500.
[18] Nawa K et al. Evaluation of a commercial automatic treatment planning system for prostate cancers. Med Dosim 2017;42:203–9.
[19] Wu B et al. Patient geometry-driven information retrieval for IMRT treatment plan quality control. Med Phys 2009;36:5497–505.
[20] Chanyavanich V, Das SK, Lee WR, Lo JY. Knowledge-based IMRT treatment planning for prostate cancer. Med Phys 2011;38:2515–22.
[21] Moore KL, Brame RS, Low DA, Mutic S. Quantitative metrics for assessing plan quality. Semin Radiat Oncol 2012;22:62–9.
[22] Fogliata A et al. On the pre-clinical validation of a commercial model-based optimisation engine: Application to volumetric modulated arc therapy for patients with lung or prostate cancer. Radiother Oncol 2014;113:385–91.
[23] Fogliata A et al. RapidPlan head and neck model: The objectives and possible clinical benefit. Radiat Oncol 2017;12.
[24] Tol JP, Delaney AR, Dahele M, Slotman BJ, Verbakel WFAR. Evaluation of a knowledge-based planning solution for head and neck cancer. Int J Radiat Oncol Biol Phys 2015;91:612–20.