Radiotherapy and Oncology 130 (2019) 75–81

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original article

Analysis of EORTC-1219-DAHANCA-29 trial plans demonstrates the potential of knowledge-based planning to provide patient-specific treatment plan quality assurance

Jim P. Tol [a], Max Dahele [a], Vincent Gregoire [b], Jens Overgaard [c], Ben J. Slotman [a], Wilko F.A.R. Verbakel [a,*]

[a] VU University Medical Center, Department of Radiation Oncology, Amsterdam, The Netherlands; [b] Université catholique de Louvain, St-Luc University Hospital, Experimental and Clinical Research Institute, Center for Molecular Imaging, Oncology and Radiotherapy, Brussels, Belgium; [c] Aarhus University Hospital, Department of Experimental Clinical Oncology, Aarhus, Denmark

ARTICLE INFO

Article history:
Received 29 December 2017
Received in revised form 29 September 2018
Accepted 1 October 2018
Available online 19 October 2018

Keywords:
Treatment planning
Quality assurance
Clinical trials
Head and neck cancer

ABSTRACT

Introduction: Radiotherapy treatment plan quality can influence clinical trial outcomes and general QA may not identify suboptimal organ-at-risk (OAR) sparing. We retrospectively performed patient-specific quality assurance (QA) of 100 head-and-neck cancer (HNC) plans from the EORTC-1219-DAHANCA-29 study.

Materials and methods: A 177-patient RapidPlan (Varian Medical Systems) model comprising institutional HNC plans was used to QA trial plans (P_{trial}). RapidPlan plans ($P_{rapidplan}$) were created using RapidPlan and Eclipse scripting to achieve a high degree of automation. Comparison between $P_{rapidplan}$ mean predicted/achieved OAR doses, and P_{trial} mean OAR doses was made for parotid/submandibular glands (PGs/SMGs) and swallowing muscles (SM).

Results: OAR predictions were made within 2 min per patient. Averaged PG/SMG/SM mean doses were 2.0/9.0/3.8 Gy lower in $P_{rapidplan}$. Using predicted $P_{rapidplan}$ combined mean OAR dose as the benchmark, a total of 60/27/4 trial plans could be improved by 3/6/9 Gy respectively.

Discussion: Individualized QA indicated that OAR sparing could frequently be improved in EORTC-1219 study plans, even though they met the trial's generic plan criteria. Automated, patient-specific QA can be performed within a few minutes and should be considered to reduce the influence of planning variation on trial outcomes.

Radiotherapy treatment plans of patients enrolled in clinical trials are commonly subject to generic quality-assurance (QA) testing to ensure that acceptable levels of organ-at-risk (OAR) sparing and planning target volume (PTV) dose coverage and homogeneity are obtained. However, passing this generic QA procedure does not necessarily indicate that the level of OAR sparing is close to optimal for any given patient. A patient-specific QA procedure is therefore desirable. This is further supported by data showing the large variation in treatment plan quality that can exist between different institutes and planners [1,2]. It has furthermore been demonstrated that sub-optimal plans may increase normal-tissue complication probability (NTCP) and reduce tumor control [3–6]. Since sub-optimal treatment plans are not considered in stratification, this may introduce bias, influence outcomes and increase the sample size necessary to detect differences between clinical trial arms.

Due to the nuances of an individual patient's geometry, it is often hard to visually determine whether (near-)optimal levels of OAR sparing have been reached. One approach involves manually replanning the treatment plans, but this is impractical, labor-intensive and prone to bias. Instead, knowledge-based planning (KBP) approaches such as RapidPlan™ (Varian Medical Systems) might be used to perform objective and automated patient-specific QA of treatment plans. RapidPlan requires the construction of a model based on a library of previously created treatment plans. This model correlates certain geometric features of the included patients (e.g. relative OAR-PTV positions and OAR/PTV volumes) with obtained OAR doses [7–9]. By analyzing the position of the OARs/PTVs on a delineated CT-scan, for each of the included OARs, the RapidPlan model performs patient-specific OAR dose predictions within a few minutes [7,10,11]. The predicted DVHs can be used to optimally position the optimization objectives thereby resulting in an actual treatment plan. Additionally, because these OAR dose predictions are in good agreement with the OAR doses in a plan made by RapidPlan, they can be used to QA a previously

* Corresponding author at: Department of Radiation Oncology, VU University Medical Center, De Boelelaan 1117, 1081HV Amsterdam, The Netherlands.
E-mail addresses: j.tol.vumc@gmail.com (J.P. Tol), w.verbakel@vumc.nl (W.F.A.R. Verbakel).

https://doi.org/10.1016/j.radonc.2018.10.005
0167-8140/

created treatment plan. The doses to all OARs can be compared with their respective predictions [12]. If the dose predictions turn out lower than the OAR doses that were achieved, the treatment plan is likely suboptimal and should be considered for replanning.

In the present work, we use this approach to create a highly automated, patient-specific plan QA workflow. We have retrospectively performed patient-specific QA of 100 plans submitted to the EORTC-1219-DAHANCA-29 study (https://clinicaltrials.gov/ct/show/NCT01880359) for locally advanced head and neck cancer (HNC).

Materials and methods

This investigation was approved by the EORTC, which made available anonymized treatment plans, consisting of delineated CT-scans and dose distributions, for 100 patients submitted from 13 centers to the EORTC-1219-DAHANCA-29-trial. These treatment plans were imported into our treatment planning system (Eclipse v15.1, Varian Medical Systems) for further analysis. Detailed study requirements can be found at http://www.eortc.org/research_field/clinical-detail/1219/. In brief, a simultaneous-integrated boost (SIB) technique is used, with boost/elective PTV (PTV_B/PTV_E) prescriptions of 70 Gy/54 Gy, delivered in 35 daily fractions of 2 Gy/1.55 Gy. Coverage and homogeneity criteria required D95%>95% and D98%>90% (Dx% = dose received by x% of structure volume), with median PTV dose ± 2% of the prescribed dose and PTV_B D5% < 107%. Intensity-modulated radiotherapy (IMRT) is mandatory (e.g. volumetric-modulated arc therapy [VMAT], TomoTherapy®). Delineated OARs always included spinal cord and brainstem (typically with 3–5 mm planning-at-risk volume [PRV] expansions), and both parotid glands; and, depending on the institute and degree of PTV overlap, mandible, submandibular glands, larynx, oral cavity, thyroid, trachea and esophagus. The planning requirements for the OARs and PTVs outlined in the trial and summarized in Table 1.

RapidPlan model creation

OAR dose predictions were generated using our institutional RapidPlan model containing 177 clinical VMAT HNC plans (Table 1) created using automatic interactive optimization for consistent, high-quality plans [13]. They had the same PTV_B/PTV_E dose pre-

Table 1
The number of structures that were included in the 100 trial plans (P_{TRIAL}) and the 177 patient head and neck cancer RapidPlan model, along with the averaged structure volumes. PTV_E was defined as the total elective PTV, with the boost PTV (PTV_B) and a 5 mm transition region subtracted. Since the RapidPlan model contained no entire larynx structure, this structure was matched to the closest corresponding structure – the upper larynx. For the mandible, only maximum dose values are considered clinically relevant. The model therefore solely assigns fixed maximum dose objectives, instead of positioning a range of optimization objectives based on the predicted doses to this structure. 'Line objectives' under the RapidPlan optimization objectives column refer to the use of a line of optimization objectives, generated by RapidPlan, used for patient-specific optimization.

	#Included in trial plans	EORTC plans Volume Range (cm^3) Mean ± StDev (Range)	#Included in RapidPlan Model	RapidPlan Model Volume Range (cm^3) Mean ± StDev (Range)	RapidPlan optimization objectives and priorities (P)	DAHANCA Trial Planning Requirements
[a]PTV_B	100	229.5 ± 121.3 (28.2–655.4)	177	181 ± 127.3 (31.1–1219)	Min. Dose: 69 Gy (2x) Max. Dose: 71 Gy P = 130	D95%[f] ≥ 95% planned dose D98% ≥ 90% planned dose D5% ≤ 107% planned dose D50% ≤ 70 Gy ± 2%
[b]PTV_E	100	367.9 ± 93.7 (70.4–654.5)	177	560.1 ± 177.4 (281.2–1930.2)	Min. Dose: 53 Gy (2x) Max. Dose: 55 Gy P = 130	D95%[f] ≥ 95% planned dose D98% ≥ 90% planned dose D50% ≤ 54.25 Gy ± 2%
Spinal Cord	100	20.8 ± 9.6 (8.9–66.1)	177	41.1 ± 17.2 (3.7–85.8)	Max. Dose: 41 Gy P = 120	Max dose (PRV) ≤ 45 Gy
Brainstem	100	26.2 ± 5.3 (11.7–36.9)	114	38.8 ± 175.8 (0.3–1568.8)	Max. Dose: 41 Gy P = 120	Max dose (PRV) ≤ 50 Gy
[c]CL Parotid	100	28.1 ± 9.9 (10.7–71.1)	177	28.3 ± 8 (4.1–49)	Line objective P = 90	Dmean[g] D ≤ 24 Gy
[d]IL Parotid	100	28.8 ± 10.4 (1.4–72)	78	28.1 ± 7.5 (11.8–47.1)	Line objective P = 90	Dmean 25–27 Gy
CL [e]SMG	60	9.6 ± 2.6 (5.2–16.6)	149	9.0 ± 2.6 (2.7–15.9)	Line objective P = 85	*Not specified*
IL [e]SMG	58	9.6 ± 2.7 (5.6–16.7)	37	9.9 ± 2.4 (6.0–14.8)	Line objective P = 85	*Not specified*
Larynx	80	49.3 ± 28.5 (0.1–120.8)	115	12.9 ± 12.9 (2.3–71.4)	Line objective P = 80	Dmean ≤ 44 Gy
Oral Cavity	91	117.5 ± 44.4 (10.4–241.1)	152	140.0 ± 72.0 (0.5–327.5)	Line objective P = 80	Dmean ≤ 30 Gy
Esophagus	44	13.8 ± 9.7 (0.1–57.4)	106	7.8 ± 9.4 (0.1–76.5)	Line objective P = 80	*Not specified*
Thyroid	24	17.8 ± 7.5 (5.2–35.5)	98	16.4 ± 11.0 (6.7–106.3)	Line objective P = 80	*Not specified*
Trachea	12	39.3 ± 12 (19.2–56.3)	107	23.6 ± 10.5 (3.8–50.8)	Line objective P = 80	*Not specified*
Mandible	96	62.5 ± 14.8 (29.3–109.1)	–	–	Max. Dose: 69 Gy P = 100	Max Dose ≤ 70 Gy

[a] PTV_B: PTV receiving the therapeutic dose of 70 Gy.
[b] PTV_E: PTV receiving the elective dose of 54 Gy.
[c] CL: contralateral.
[d] IL: ipsilateral.
[e] SMG: submandibular gland.
[f] Dx% = Dose received by x% of the volume.
[g] Dmean = mean dose.

scription as the trial and attempted, where feasible, to spare the oral cavity, parotid and submandibular glands, and various structures in the neck (e.g. individual swallowing structures [14], lower/upper larynx, esophagus, trachea and thyroid). The optimization objectives used by RapidPlan are shown in Table 1.

Automated plan QA

Using the Eclipse™ application programming interface (API, Varian Medical Systems v15.1), scripts were written to fully automate OAR dose predictions and subsequent creation of dual-arc VMAT plans solely using a delineated CT-scan as input. The delineations provided by the EORTC were used, and no additional OAR contours were created. The software allowed the work to be batched and took into account variation in structure nomenclature. A detailed explanation of the automated treatment planning process is provided in the supplementary materials. In brief, the software automatically positions dual-arc VMAT fields. The RapidPlan model is used to predict the range of achievable DVHs for all OARs (e.g. salivary glands, swallowing muscles and oral cavity), and position minimum/maximum dose objectives for the boost/elective PTVs, and maximum dose objectives for the spinal cord, brainstem and mandible. Lines of optimization objectives are automatically positioned along the inferior boundary of the DVH-prediction range of the parallel OARs, allowing for an automated optimization process, which is performed subsequently by the script, resulting in an actual treatment plan after performing dose calculation (Acuros v13.7). In cases where the EORTC plan used IMRT, we nonetheless made a VMAT plan, although the same optimization objectives could have been used to drive IMRT optimization.

Study endpoints

After running the script, two treatment plans were available for each patient:

1. A trial plan (P_{trial}) from a center participating in the DAHANCA trial, with associated OAR and PTV metrics.
2. A RapidPlan plan ($P_{rapidplan}$) providing predicted and achieved DVHs for all OARs that were delineated in P_{trial}.

To facilitate OAR dose comparisons between P_{trial} and the RapidPlan predictions, a "predicted mean dose" was constructed for each OAR by firstly calculating the 'mid-prediction' DVH by averaging the doses at of the upper and lower DVH prediction boundary at each volume percentage, and secondly calculating the mean dose of this mid-prediction DVH. To conform to previous findings [13], the accuracy of the OAR dose predictions was evaluated by comparing the predicted and achieved mean OAR doses in $P_{rapidplan}$, along with the fraction of the achieved DVH-lines falling below/inside/above the DVH prediction ranges.

Resulting $P_{rapidplan}$ and P_{trial} plan quality was assessed using protocol PTV_B/PTV_E dose coverage criteria (e.g. doses received by 95%/98% of the PTV volume [D95%/D98%], and D5% and the homogeneity index [HI = 100% * (D2%–D98%)/D50%] for PTV_B. $P_{rapidplan}$ was normalized such that the mean PTV_B doses were equal to P_{trial}, which were exported including normalization. Both sets of plans were compensated for PTVs approaching the surface of the body; $P_{rapidplan}$ included a virtual bolus region, while P_{trial}, depending on the center that provided the plan, could either use a virtual-bolus region, or have PTVs cropped underneath the skin. For the analysis between Ptrial and Prapidplan, we used the same PTV structures.

Furthermore, a comparison was made between maximum mandible, spinal cord, brainstem doses; mean OAR doses including composite dose-reporting structures containing all salivary/swallowing structures ($comp_{sal}/comp_{swal}$); and normal-tissue doses (e.g. V5Gy/V30Gy/V50Gy/mean dose to body-PTV). Differences were assessed using a paired, two-sided Student *t*-test, with $p < 0.05$ considered significant. Time requirements for the various aspects of the treatment planning process were determined for the first 5 patients. For illustrative purposes, it was evaluated for how many OARs the mean doses differed more than 3/6/9 Gy between the RapidPlan predicted OAR doses, and the $P_{rapidplan}$ and P_{trial} achieved OAR doses.

Results

Averaged over five patients, automated positioning of the treatment fields in $P_{rapidplan}$ and generating the DVH-predictions and patient-specific optimization objectives required 1.7 ± 0.2 minutes, VMAT optimization and dose calculation 15.3 ± 2.2 minutes, and the final plan was completed within 28.4 ± 3.9 minutes.

Table 2 summarizes the dosimetric results, averaged over all patients. One P_{trial} and one $P_{rapidplan}$ plan, of different patients, failed PTV dose coverage and homogeneity criteria, violating the D98% < 90% requirement of PTV_E. Although all remaining RapidPlan plans fulfilled the trial criteria, averaged over all patients, P_{trial} PTV_B dose coverage and homogeneity values were significantly better. PTV_E dose coverage and homogeneity, and maximum doses to spinal cord, brainstem and mandible were similar with a maximum difference of only 1.8 Gy between P_{trial} and $P_{rapidplan}$. Averaged over all plans, $P_{rapidplan}$ achieved significantly lower mean doses for all OARs (e.g. 1.5/9.1 Gy for CL parotid/submandibular glands; and 3.1/6.3 Gy for larynx/esophagus). In addition, $P_{rapidplan}$ Body-PTV mean dose, V5 Gy and V30 Gy decreased by 1.7 Gy, 2.5% and 0.6%, respectively, while V50 Gy increased by 1.6%.

There was close correspondence between the predicted and achieved mean OAR doses in $P_{rapidplan}$ (Table 2), with average differences of 0.8 ± 0.7 Gy (0.1–2.5 Gy). The high prediction accuracy of the model is also demonstrated by Fig. 1. For the majority of OARs (442/570), the $P_{rapidplan}$ achieved mean dose is within ±3 Gy of predicted. In contrast, P_{trial} mean doses were considerably higher than the Rapidplan predicted doses (Fig. 1B) and the mean dose could be improved by >3 Gy for 293/570 OARs, while respectively 187/570 and 116/570 OARs could be spared by >6 Gy and >9 Gy.

Table S1 (Supplementary materials) shows the DVH-prediction width for each OAR, averaged over all patients. The DVH-prediction bands were generally smallest for the CL parotid (5.4 ± 0.5 Gy), and largest for IL submandibular gland (12.6 ± 1.5 Gy). This Table also shows the percentage of DVH points falling within, above and below the DVH prediction range, averaged over all patients. For $P_{rapidplan}$, 77.8 ± 13.2% of the achieved DVH-lines was located within the prediction range, while for P_{trial} this was 43.5 ± 14.8%.

The histogram in Fig. 2 illustrates the resulting dose differences between the achieved mean dose in P_{trial} and the predicted mean dose in $P_{rapidplan}$. To balance the dose contributions of similar OARs, mean dose differences were averaged over both parotid glands, both submandibular glands, and over all swallowing structures. These data show that for, the majority of structures, the predicted mean dose was substantially lower than the mean dose achieved by the trial plan. This indicates, without making a treatment plan, that a high proportion of P_{trial} plans could be improved. These differences were especially prominent for the swallowing muscles and submandibular glands, of which respectively 39 (out of 93) and 43 (out of 61) could have been spared by >6 Gy. Conversely, oral cavity mean doses were occasionally lower in P_{trial} than predicted by RapidPlan, with dose differences of 3 (out of 92) of these structures exceeding 9 Gy. For only 11/100 patients, achieved mean OAR doses, averaged over the structures shown in Fig. 2, were 0–3 Gy lower in P_{trial} than predicted by RapidPlan. Conversely, in 33/100, 23/100 and 4/100 plans, predicted mean doses

Table 2
Dosimetric parameters of the planning target volumes (PTVs) and organs-at-risk (OARs) found in the trial and RapidPlan plans (P_{TRIAL} and $P_{RAPIDPLAN}$, respectively), averaged over all included patients (n = 100). Results show mean dose ± standard deviation.

Dosimetric Parameter	Structure	P_{TRIAL} Achieved	$P_{RAPIDPLAN}$ Achieved	$P_{RAPIDPLAN}$ Predicted
[f]D95% (%)	Boost PTV	96.4 ± 1.1	95.9 ± 0.8*	
D98% (%)		94.8 ± 1.8	94.1 ± 1.6*	
D5% (%)		102.6 ± 0.8	103.1 ± 0.6*	
HI (%)		8.4 ± 2.4	9.7 ± 2.3*	
D95% (%)	Elective PTV	98.3 ± 4.4	97.7 ± 0.7	
D98% (%)		96.1 ± 5.2	96.1 ± 1.1	
[g]V95% (%)		98.5 ± 1.4	98.9 ± 0.8*	
MaxDose (Gy)	Spinal Cord	37.5 ± 6.8	39.1 ± 4.1*	
	Spinal Cord PRV	43.8 ± 4.7	44.4 ± 2.2	
	Brainstem	33.2 ± 12.4	31.4 ± 11.9*	
	Brainstem PRV	38.7 ± 12.7	36.9 ± 13.3*	
	Mandible	67.5 ± 7.8	67.3 ± 8.5	
Mean Dose (Gy)	[a]CL Parotid	23.3 ± 5.6	21.8 ± 6.9*	21.7 ± 5.6
	[b]IL Parotid	29.5 ± 10.3	27.0 ± 8.6*	27.1 ± 6.9
	CL [c]SMG	50.7 ± 11.7	41.6 ± 12.8*	39.1 ± 8.4
	IL [c]SMG	60.5 ± 9.3	55.6 ± 10.6*	54.5 ± 7.1
	[d]$Comp_{sal}$	31.2 ± 6.6	28.3 ± 6.9*	
	Larynx	51.4 ± 15.4	48.3 ± 17.0*	47.6 ± 15.2
	Oral Cavity	32.9 ± 12.5	30.1 ± 12.3*	31.0 ± 10.8
	Thyroid	50.4 ± 8.3	48.4 ± 9.4*	47.9 ± 7.7
	Trachea	34.3 ± 9.8	29.7 ± 9.0*	29.9 ± 8.3
	Esophagus	31.7 ± 13.9	25.4 ± 14.2*	24.6 ± 13.0
	[e]$Comp_{swal}$	47.0 ± 13.6	43.2 ± 15.0*	
Body	Mean Dose (Gy)	27.1 ± 2.8	25.4 ± 2.5*	
	V5Gy (%)	91.3 ± 5.5	88.8 ± 2.7*	
	V30Gy (%)	35.7 ± 6.4	35.1 ± 5.9	
	V50Gy (%)	14.1 ± 3.8	15.8 ± 3.0*	

[a] CL: contralateral.
[b] IL: ipsilateral.
[c] SMG: submandibular gland.
[d] $Comp_{sal}$: composite salivary glands.
[e] $Comp_{swal}$: composite swallowing muscles.
[f] Dx%: Dose received by x% of the PTV volume, relative to PTV prescription dose.
[g] VxGy: volume receiving x% of the prescribed dose.
* Indicates a significant difference with 'P_{TRIAL} Achieved'.

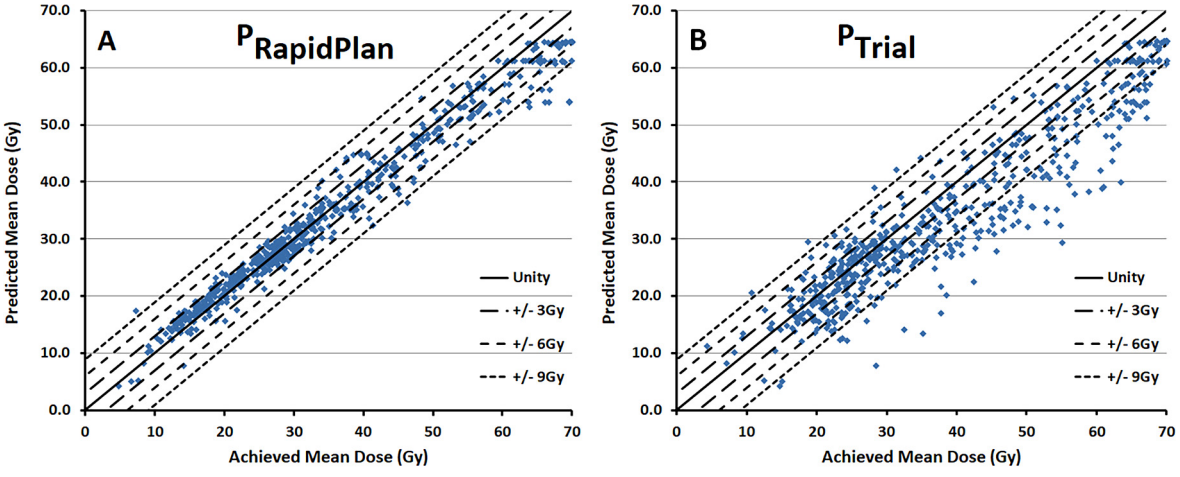

Fig. 1. The mean dose of the DVH prediction range (y-axis), plotted against the mean dose achieved in $P_{rapidplan}$ (Fig. A) and P_{trial} (Fig. B) for all OARs (n = 570). The unity is indicated with the solid lines, while the long dashed, short dashed and dotted lines represent ±3 Gy, ±6Gy and ±9 Gy dose differences between the predicted and achieved mean OAR doses.

were respectively 3–6 Gy, 6–9 Gy and >9 Gy lower in $P_{rapidplan}$ than achieved in P_{trial}.

Discussion

In this proof of principle analysis, we created a highly automated patient-specific plan QA workflow using a commercial KBP solution and software developed using the Eclipse application programming interface. Performing patient-specific QA, in contrast to evaluating plans using generic QA criteria, ensures that only high-quality plans are accepted into clinical trials. As a result, differences in outcome are less likely to be influenced by variations in treatment plan quality, and more likely to be due to the intervention under investigation, potentially increasing statistical power. The flowchart in Fig. 3 suggests a possible workflow for applying

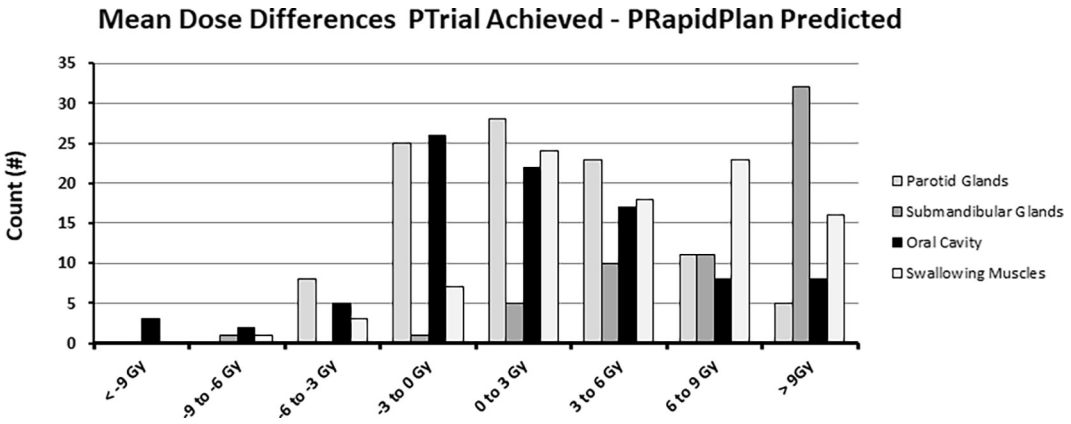

Fig. 2. A Histogram represents the differences between P_{trial} achieved mean dose and the predicted mean dose in $P_{rapidplan}$ for the salivary glands, swallowing muscles, oral cavity and the average of these structures.

OAR DVH predictions generated by RapidPlan to provide patient-specific QA of plans. Using DVH-predictions alone identified 60/100 trial plans for which average OAR doses could potentially be improved by >3 Gy. Note that this threshold was chosen purely arbitrarily to demonstrate patient-specific plan QA approach using RapidPlan. Within each trial it could be discussed which deviation of OAR doses from the prediction would be acceptable. The largest improvements were often noted in patients with only a limited number of delineated OARs (Table 1). An important pre-requisite for applying RapidPlan in this fashion is high prediction accuracy of the model, i.e., a close correspondence between the DVH-lines that were predicted by RapidPlan, and the DVH-lines that were achieved when creating the actual treatment plan. For our model, this was demonstrated by Fig. 1, Table 2 and Table S1. Since treatment plans that comprised the RapidPlan model take into account all delineated structures when performing the DVH-predictions, it is possible that the amount of achievable OAR sparing was underestimated in patients with a limited number of delineated OARs, meaning that further improvements in sparing might be feasible.

Although the absolute differences were small, PTV_B dose coverage and homogeneity values were significantly worse in $P_{rapidplan}$ compared to P_{trial}. This is likely to have several causes, including differences in plan normalization, or PTV optimization priorities being higher in some of the trial plans. Because our aim was to perform this analysis with a workflow that was as automated as possible, it was not attempted to manually re-optimize individual plans or super-impose DVH curves. The risk of this approach is that, since there is an exponential trade-off between OAR sparing and PTV dose homogeneity [15], this may have resulted in an over-estimation of the potential for additional OAR sparing. We therefore performed a secondary analysis to see if this was likely. Table 3 shows data for those 50/100 patients with the PTV_B homogeneity index of $P_{rapidplan}$ that was at a maximum 1% worse than achieved in $P_{trial.}$ These results show similar levels of OAR sparing as demonstrated in Table 2, indicating that the improvements in predicted and achieved OAR sparing in $P_{rapidplan}$ were not driven primarily by the differences in PTV dose coverage and homogeneity. In addition, a relatively large sample size (100 plans) was used to try and improve the robustness of the conclusions. We have not specifically evaluated the extent to which differences in approach to normalization; differences in the way in which the plans handled PTV dose coverage when the PTV approached the skin (e.g. PTV cropping vs bolus); and differences between planning systems (including in performance close to the surface region) may have influenced PTV dosimetry. Additional attention is needed to such issues if fully automated QA approaches were to be implemented in practice: how to handle PTV structures in which surface compensation methods and normalization differ? Should you benchmark OAR sparing in the submitted plan against a plan providing the same PTV dosimetry, or one that provides a little less PTV dose, but still meets the trial criteria? Additional challenges of practical importance for automation were encountered, including varying nomenclature and spelling of structures and multiple structures with similar names.

Normal-tissue complication probability (NTCP) models can be used to assess the impact of suboptimal treatment plans. For example, for 16/100 plans, the average predicted doses to the parotid glands were >6 Gy lower than the mean dose achieved in the

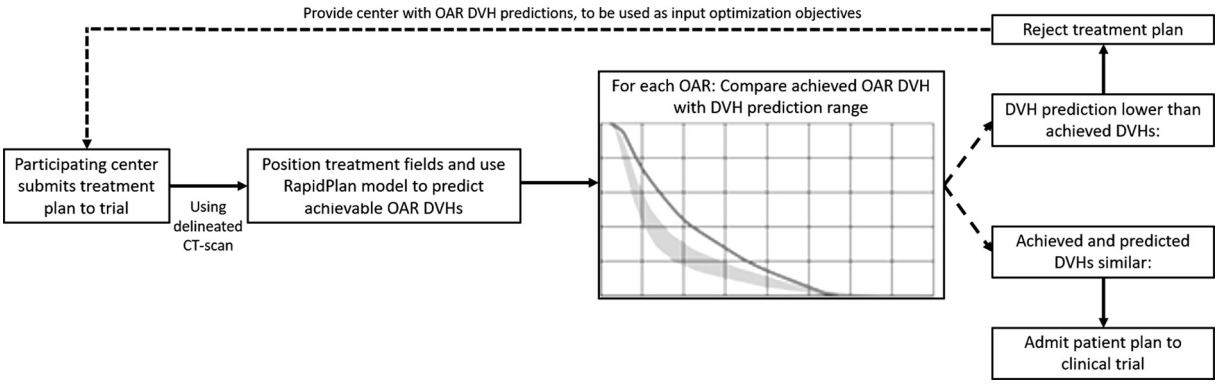

Fig. 3. A flowchart proposing how a clinical trial can use the organ-at-risk (OAR) dose–volume histogram (DVH) predictions generated by a RapidPlan model to perform patient-specific quality assurance of submitted treatment plans.

Table 3
Dosimetric data for the subset of 50/100 patients for which the $P_{rapidplan}$ PTV_B homogeneity index was at maximum 1% worse than achieved in P_{trial}. These results demonstrate the improved levels of OAR sparing in $P_{rapidplan}$ did not result from the slight decreases in PTV dose coverage and homogeneity compared to the trial plans. For reference, the dosimetric values obtained over the entire cohort (n = 100) of patients are shown next to the subset of patients.

Dosimetric parameter	Structure	P_{TRIAL} achieved (n = 50/100)	$P_{RAPIDPLAN}$ achieved (n = 50/100)	P_{TRIAL} achieved (n = 100)	$P_{RAPIDPLAN}$ achieved (n = 100)
[f]D95% (%)	Boost PTV	95.8 ± 0.9	96.2 ± 0.7*	96.4 ± 1.1	95.9 ± 0.8
D98% (%)		93.9 ± 1.7	94.7 ± 0.8*	94.8 ± 1.8	94.1 ± 1.6
D5% (%)		103.1 ± 0.7	102.9 ± 0.4	102.6 ± 0.8	103.1 ± 0.6
HI (%)		9.9 ± 2	8.9 ± 1.3*	8.4 ± 2.4	9.7 ± 2.3
D95% (%)	Elective PTV	97.2 ± 1.7	97.9 ± 0.4*	98.3 ± 4.4	97.7 ± 0.7
D98% (%)		94.4 ± 3.5	96.4 ± 0.5*	96.1 ± 5.2	96.1 ± 1.1
[g]V95% (%)		97.8 ± 1.5	99.1 ± 0.4*	98.5 ± 1.4	98.9 ± 0.8
MaxDose (Gy)	Spinal Cord	37.3 ± 9.2	37.9 ± 7.9	37.5 ± 6.8	39.1 ± 4.1
	Spinal Cord PRV	44.1 ± 3.3	44.6 ± 2.4	43.8 ± 4.7	44.4 ± 2.2
	Brainstem	33.8 ± 13.2	31.6 ± 12.1*	33.2 ± 12.4	31.4 ± 11.9
	Brainstem PRV	38.7 ± 13.8	37.0 ± 14.1	38.7 ± 12.7	36.9 ± 13.3
	Mandible	67.1 ± 12.5	66.8 ± 12.8	67.5 ± 7.8	67.3 ± 8.5
Mean Dose (Gy)	[a]CL Parotid	23.2 ± 5.4	22.4 ± 6.9	23.3 ± 5.6	21.8 ± 6.9
	[b]IL Parotid	29.5 ± 10.7	27.8 ± 9.3*	29.5 ± 10.3	27.0 ± 8.6
	CL [c]SMG	50.1 ± 11.9	40.5 ± 12.2*	50.7 ± 11.7	41.6 ± 12.8
	IL [c]SMG	60.3 ± 10.1	55.8 ± 10.2*	60.5 ± 9.3	55.6 ± 10.6
	[d]$Comp_{sal}$	49.3 ± 15.7	46.8 ± 16.8*	31.2 ± 6.6	28.3 ± 6.9
	Larynx	33.6 ± 12.4	31.4 ± 12.1*	51.4 ± 15.4	48.3 ± 17.0
	Oral Cavity	51.6 ± 10.4	50.0 ± 11.0*	32.9 ± 12.5	30.1 ± 12.3
	Thyroid	33.0 ± 11.6	29.7 ± 12.0*	50.4 ± 8.3	48.4 ± 9.4
	Trachea	30.4 ± 13.5	24.6 ± 13.0*	34.3 ± 9.8	29.7 ± 9.0
	Esophagus	31.6 ± 6.7	29.2 ± 6.8*	31.7 ± 13.9	25.4 ± 14.2
	[e]$Comp_{swal}$	45.3 ± 13.9	42.5 ± 14.9*	47.0 ± 13.6	43.2 ± 15.0

[a] CL: contralateral.
[b] IL: ipsilateral.
[c] SMG: submandibular gland.
[d] $Comp_{sal}$: composite salivary glands.
[e] $Comp_{swal}$: composite swallowing muscles.
[f] Dx%: Dose received by x% of the PTV volume, relative to PTV prescription dose.
[g] VxGy: volume receiving x% of the prescribed dose.
* Indicates a significant difference with 'P_{TRIAL} Achieved (n = 50/100)'.

submitted plan (P_{trial}). Evaluating these dose differences using the NTCP model of Dijkema et al.[3] shows that, depending on the initial dose, the NTCP for a >75% reduction in salivary flow ratio after treatment, may decrease by >10%. At the least, the present results suggest the need to consider plan quality variations in each clinical trial arm when stratifying patients and interpreting outcomes.

Assessment of the generic QA process in this trial based on a benchmark case sent to participating centers before patient inclusion concluded that "the overall results of the treatment plans were satisfactory, without the presence of unacceptable protocol variations" [16]. However, the present analysis shows the added benefits of patient-specific QA. In addition, the fact that we observed substantial variations in plan quality even when all trial plans were made using techniques available for many years, indicates the scope for additional training/improvement strategies in advanced planning within the radiotherapy community. This is illustrated by, for example, the results for the submandibular glands, which in 43 plans could have been spared by >6 Gy.

Moore et al. used an in-house KBP solution for patient-specific plan benchmarking in the RTOG 0126 prostate cancer trial, and identified a substantial number of patients at risk for increased rectal toxicity due to suboptimal plans [17]. In the present study, we analyzed a complex tumor site and used the combination of a commercial KBP solution and scripting to develop a process allowing for near real-time patient-specific QA. A further strength of the current study is that the RapidPlan model was created using automatically optimized treatment plans made with a consistent planning strategy [13]. If a submitted plan is deemed suboptimal after patient-specific QA, the trial committee could choose for example to send the predicted OAR DVHs to the submitting center who may decide to use this to improve their optimization. Such steps also have the potential to be largely automated.

Strong correlation was observed between predicted and achieved RapidPlan OAR doses despite the fact that the model was constructed using clinical treatment plans from a single institute, rather than trial plans, an approach used by Geng et al. [18] and even though the trial plans varied considerably in the number of delineated OARs (Table 1), OAR contouring, optimization strategies and treatment techniques. However, it should be noted that the 3 oral cavities were spared >9 Gy better than predicted by our RapidPlan model (Fig. 2), are likely the result of contouring differences, which may be especially prominent for this structure as it is not a well-defined organ. Such differences may have caused extrapolation errors by our model, resulting in relatively poor predictions. Variation in salivary glands and swallowing structures delineations have been shown not to lead to inferior dose predictions [19].

One patient was identified for which $P_{rapidplan}$ was worse than P_{trial}. Detailed analysis indicated that this was likely due to the PTV_B volume being considerably larger than the average in the model (486.0 cm^3 versus 181.0 ± 127.3 cm^3), while PTV_E was smaller (259.8 cm^3 versus 560.1 ± 177.4 cm^3). This led to poor DVH predictions and suboptimal OAR sparing. Strategies to overcome this could include increasing the size of the RapidPlan model to add more variation in PTV-OAR geometries and to incorporate limits above which a KBP may be deemed not suitable for an individual patient.

A number of other limitations deserve mention. In our model, the swallowing muscles were contoured individually (upper/lower larynx, inferior/medial/superior pharyngeal constrictor muscle), whereas only the combined structures were available in the EORTC plans. Since the volume of these structures was therefore typically larger in the EORTC plans, the model had to perform extrapolations, potentially degrading the OAR predictions. Such volumetric

differences might have affected more structures, considering for example, the large differences in PTV_E size between P_{trial} and $P_{RapidPlan}$. Regardless of these volumetric differences, RapidPlan could still be successfully applied to patients contoured following different delineation strategies, which can be considered a strength of the knowledge-based planning approach to provide automated QA of clinical trial plans. It is also important to note that our proposed automated plan QA approach is highly depended on the quality of the KBP model library. If poor quality treatment plans had been used to construct the model, the DVH-prediction ranges would become wider, leading to the rejection of fewer patients. There is currently no standard way to objectively determine the quality of plans that might be included in a model. Finally, our RapidPlan model was constructed solely using VMAT plans. The resulting, VMAT-shaped DVH-predictions, were not only used to benchmark P_{trial} plans created using VMAT, but also included IMRT and TomoTherapy plans. This difference in planning/delivery technique may have had some influence on the percentage of the achieved DVH-curves falling below/inside/above the DVH-prediction, as presented in Table S1. However, we believe this effect is likely to have been modest, since similar levels of OAR sparing can be obtained between these techniques [20,21]. It should also be noted that our institutional plans that comprised the model, and the plans submitted to the DAHANCA trial, may have differed in terms of planning aims. However, the planning aims of the trial (shown in Table 1) are the minimum requirements that should be fulfilled. Treating centers often use additional planning requirements that could be stricter than those mentioned in the trial. For example, the PTV coverage requirement in the trial is D95% > 95%, whereas our institutional protocol requires D98 > 95%. Using a model that satisfied the trial criteria regarding PTV dose coverage and homogeneity would therefore likely have led to lower OAR dose predictions. Finally, the specific treatment planning system (Eclipse) and optimization/dose calculation algorithms used may influence the results.

In conclusion, the present work has demonstrated the role of knowledge-based planning to provide largely automated patient-specific QA of treatment plans using individualized model-based dose predictions. It has also highlighted that there is room for improvement in how advanced treatment plans are made. When used on HNC treatment plans that were submitted to an ongoing trial, improvements of >3 Gy in OAR sparing could be obtained in about three-quarters of patients.

Conflicts of interest

The Department of Radiation Oncology of the VU University Medical Center has research agreements with Varian Medical Systems.

MD, BS, and WV have received speaker honoraria from Varian Medical Systems.

Acknowledgements

We kindly acknowledge Enrico Clementel for facilitating the data transfer and assisting with various inquiries.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2018.10.005.

References

[1] Nelms BE, Robinson G, Markham J, Velasco K, Boyd S, Narayan S, et al. Variation in external beam treatment plan quality: an inter-institutional study of planners and planning systems. Pract Radiat Oncol 2012;2:296–305. https://doi.org/10.1016/j.prro.2011.11.012.

[2] Chanyavanich V, Das SK, Lee WR, Lo JY. Knowledge-based IMRT treatment planning for prostate cancer. Med Phys 2011;38:2515. https://doi.org/10.1118/1.3574874.

[3] Dijkema T, Raaijmakers CPJ, Ten Haken RK, Roesink JM, Braam PM, Houweling AC, et al. Parotid gland function after radiotherapy: the combined michigan and utrecht experience. Int J Radiat Oncol Biol Phys 2010;78:449–53. https://doi.org/10.1016/j.ijrobp.2009.07.1708.

[4] Murdoch-Kinch C-A, Kim HM, Vineberg KA, Ship JA, Eisbruch A. Dose-effect relationships for the submandibular salivary glands and implications for their sparing by intensity modulated radiotherapy. Int J Radiat Oncol Biol Phys 2008;72:373–82. https://doi.org/10.1016/j.ijrobp.2007.12.033.

[5] Christianen MEMC, Schilstra C, Beetz I, Muijs CT, Chouvalova O, Burlage FR, et al. Predictive modelling for swallowing dysfunction after primary (chemo) radiation: results of a prospective observational study. Radiother Oncol 2012;105:107–14. https://doi.org/10.1016/j.radonc.2011.08.009.

[6] Beetz I, Schilstra C, van der Schaaf A, van den Heuvel ER, Doornaert P, van Luijk P, et al. NTCP models for patient-rated xerostomia and sticky saliva after treatment with intensity modulated radiotherapy for head and neck cancer: the role of dosimetric and clinical factors. Radiother Oncol 2012;105:101–6. https://doi.org/10.1016/j.radonc.2012.03.004.

[7] 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. https://doi.org/10.1016/j.ijrobp.2014.11.014.

[8] Fogliata A, Wang P-M, Belosi F, Clivio A, Nicolini G, Vanetti E, et al. Assessment of a model based optimization engine for volumetric modulated arc therapy for patients with advanced hepatocellular cancer. Radiat Oncol 2014;9:236. https://doi.org/10.1186/s13014-014-0236-0.

[9] Alongi F, Bignardi M, Garassino I, Pentimalli S, Cavina R, Mancosu P, et al. Prospective phase II trial of cetuximab plus VMAT-SIB in locally advanced head and neck squamous cell carcinoma. Feasibility and tolerability in elderly and chemotherapy-ineligible patients. Strahlenther Onkol 2012;188:49–55. https://doi.org/10.1007/s00066-011-0006-y.

[10] Yuan L, Ge Y, Lee WR, Yin FF, Kirkpatrick JP, Wu QJ. Quantitative analysis of the factors which affect the interpatient organ-at-risk dose sparing variation in IMRT plans. Med Phys 2012;39:6868–78. https://doi.org/10.1118/1.4757927.

[11] 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. https://doi.org/10.1118/1.4761864.

[12] Tol JP, Dahele M, Delaney AR, Slotman BJ, Verbakel WFAR. Can knowledge-based DVH predictions be used for automated, individualized quality assurance of radiotherapy treatment plans? Radiat Oncol 2015;10:234. https://doi.org/10.1186/s13014-015-0542-1.

[13] Tol JP, Dahele M, Peltola J, Nord J, Slotman BJ, Verbakel WF. Automatic interactive optimization for volumetric modulated arc therapy planning. Radiat Oncol 2015;10:75. https://doi.org/10.1186/s13014-015-0388-6.

[14] Christianen MEMC, Langendijk JA, Westerlaan HE, van de Water TA, Bijl HP. Delineation of organs at risk involved in swallowing for radiotherapy treatment planning. Radiother Oncol 2011;101:394–402. https://doi.org/10.1016/j.radonc.2011.05.015.

[15] Tol JP, Dahele M, Doornaert P, Slotman BJ, Verbakel WFAR. Toward optimal organ at risk sparing in complex volumetric modulated arc therapy: an exponential trade-off with target volume dose homogeneity. Med Phys 2014;41. https://doi.org/10.1118/1.4862521. 021722.

[16] Christiaens M, Collette S, Overgaard J, Gregoire V, Kazmierska J, Castadot P, et al. Quality assurance of radiotherapy in the ongoing EORTC 1219-DAHANCA-29 trial for HPV/p16 negative squamous cell carcinoma of the head and neck: results of the benchmark case procedure. Radiother Oncol 2017;123:424–30. https://doi.org/10.1016/j.radonc.2017.04.019.

[17] Moore KL, Schmidt R, Moiseenko V, Olsen LA, Tan J, Xiao Y, et al. Quantifying unnecessary normal tissue complication risks due to suboptimal planning: a secondary study of RTOG 0126. Int J Radiat Oncol Biol Phys 2015;92:228–35. https://doi.org/10.1016/j.ijrobp.2015.01.046.

[18] Geng H, Giaddui TG, Zhong H, Rosenthal DI, Galvin JM, Xiao Y, et al. Knowledge engineering-based quality evaluation of NRG oncology RTOG 0522 treatment plans. Int J Radiat Oncol 2017;99:S174. https://doi.org/10.1016/j.ijrobp.2017.06.439.

[19] Delaney AR, Dahele M, Slotman BJ, Verbakel WFAR. Is accurate contouring of salivary and swallowing structures necessary to spare them in head and neck VMAT plans. Radiother Oncol 2018.

[20] Stieler F, Wolff D, Schmid H, Welzel G, Wenz F, Lohr F. A comparison of several modulated radiotherapy techniques for head and neck cancer and dosimetric validation of VMAT. Radiother Oncol 2018;101:388–93.

[21] van Gestel D, van Vliet-Vroegindeweij C, van den Heuvel F, Crijns W, Coelmont A, et al. RapidArc, SmartArc and TomoHD compared with classical step and shoot and sliding window intensity modulated radiotherapy in an oropharyngeal cancer treatment plan comparison. Radiat Oncol 2013;8:37.