Automated proton therapy planning: clinical validation for head-and-neck cancer and chordoma

Huiskes, M.

Chapter 3

Validation of Fully Automated Robust Multi-criterial Treatment Planning for Head and Neck Cancer IMPT

Huiskes, M., Kong, W., Oud, M., Crama, K., Rasch, C., Breedveld, S., Heijmen, B., & Astreinidou, E.

International Journal of Radiation Oncology, Biology, Physics (2024)
https://doi.org/10.1016/J.IJROBP.2023.12.034

ABSTRACT

Purpose: To compare robust intensity modulated proton therapy (IMPT) plans, automatically generated with wish-list based multi-criterial optimization as implemented in Erasmus-iCycle, with manually created robust clinical IMPT plans, for patients with head and neck cancer.

Methods and Materials: Thirty-three patients with head and neck cancer were retrospectively included. All patients were previously treated with a manually created IMPT plan with 7000 cGy dose prescription to the primary tumor (CTV7000) and 5425 cGy dose prescription to the bilateral elective volumes (CTV5425). Plans had a four-beam field configuration and were generated with scenario-based robust optimization (21 scenarios, 3 mm setup error and ±3% density uncertainty for the CTVs). Three clinical plans were used to configure the Erasmus-iCycle wish-list for automated generation of robust IMPT plans for the other 30 included patients, in line with clinical planning requirements. Automatically and manually generated IMPT plans were compared for (robust) target coverage, organ-at-risk (OAR) doses and normal tissue complication probabilities (NTCP). No manual fine-tuning of automatically generated plans was performed.

Results: For all automatically generated plans, voxel-wise minimum D98% values for the CTVs were within clinical constraints and similar to manual plans. All investigated OAR parameters were favorable in the automatically generated plans (all $p<0.001$). Median reductions in mean dose to OARs went up to 667 cGy for the inferior pharyngeal constrictor muscle, while median reductions in D0.03 cm^3 in serial OARs ranged up to 1795 cGy for the spinal cord surface. The observed lower mean dose in parallel OARs resulted in statistically significant lower NTCP for xerostomia (grade ≥2: 34.4% vs. 38.0%, grade ≥3: 9.0% vs. 10.2%) and dysphagia (grade ≥2: 11.8% vs. 15.0%, grade ≥3: 1.8% vs. 2.8%).

Conclusions: Erasmus-iCycle was able to produce IMPT dose distributions fully automatically with similar (robust) target coverage and improved OAR doses and NTCPs as compared to clinical manual planning, with negligible hands-on planning workload.

INTRODUCTION

Radiation therapy (RT) has a key role in the curative treatment of head and neck cancer (HNC) patients, aiming to control tumors while minimizing radiation-induced damage to the surrounding organs at risk (OARs). Intensity modulated photon radiotherapy (IMRT) and volumetric arc therapy (VMAT) are standard clinical practice in radiation therapy.[1] Since several years, intensity modulated proton therapy (IMPT) has become more widely available as an alternative treatment technique.[2]

Proton beams have a sharp distal dose fall-off (Bragg peak), resulting in minimal to no exit dose and lower entrance dose compared to photon beams, both contributing to reduced doses in surrounding healthy tissues while preserving conformal target coverage.[3] However, proton therapy is sensitive to patient set-up uncertainties, tissue density inhomogeneities, and anatomical changes. To prevent unacceptable deviations of delivered doses from planned doses, robust optimization can be applied [4], where multiple scenarios with different density- and setup settings are taken into account during optimization.[5]

For HNC patients with multiple targets with complex shapes and different dose levels, situated in close proximity of several OARs, finding optimal dose trade-offs between the targets and all OARs with interactive manual treatment planning is complex, requiring a high level of planner intervention. This makes the manual treatment planning process workload intensive, time consuming and may delay the start of the treatment. In addition, the resulting plan quality may highly depend on the experience and time investment of the planner.[6,7] Automated treatment planning could be a possible solution for this, as it can drastically reduce the hands-on planning-time, while potentially also gain in plan quality and consistency.

Over the past years, many in-house developed or commercially available automated planning approaches were proposed, as described in the review by Hussein et al. [8]. With knowledge based- or machine learning planning, a library of previously treated patient- and plan parameters are used to predict a new plan [9,10], whereas in multi-criteria optimization (MCO), generated plans are Pareto optimal. In *a posteriori* MCO, a set of Pareto optimal plans is automatically generated for each patient and the final, clinically favorable plan is selected by a planner with so-called Pareto navigation [11]. In *a priori* MCO, a single plan is fully automatically generated for each patient; this plan is Pareto optimal and clinically favorable. Breedveld et al. [12] from Erasmus MC in Rotterdam have developed Erasmus-iCycle for *a priori* MCO. With Erasmus-iCycle, each fully automated multi-criterial

plan generation is steered by a treatment-site specific wish-list, containing hard planning constraints and prioritized planning objectives. During plan generation, the objectives from the wish-list are sequentially minimized in the order of the given priorities, while avoiding violation of the constraints. After each objective minimization a constraint is added to the problem, such that its value is maintained in subsequent minimizations of lower priority objectives.[12,13]

Validation studies for automated planning methods have shown successes for different cancer types, including HNC.[10,14–16] While most studies have been performed for photon beam therapy, studies for proton therapy are scarce. Validations of Erasmus-iCycle automated planning in photon therapy have consistently demonstrated enhanced plan quality compared to manual planning for various treatment sites, including HNC.[17–25] Since 2012, Erasmus-iCycle has been in clinical use for IMRT and VMAT at Erasmus MC. Erasmus-iCycle was extended with proton pencil beam scanning and further developments for IMPT optimization were implemented.[26–28] Automated IMPT plan generation with Erasmus-iCycle has already been used in several studies.[29–31] In addition, Erasmus-iCycle was investigated for preselection of patients for final selection for IMPT with model-based photon-proton plan comparisons, as implemented in the Netherlands.[32] However, clinical validations of Erasmus-iCycle in IMPT are still lacking.

The aim of this study was to dosimetrically compare robust IMPT plans, fully automatically generated with Erasmus-iCycle, with clinically delivered robust IMPT plans that were manually generated with our clinical treatment planning system (RayStation) for HNC patients.

METHODS AND MATERIALS

Patients

A total of 33 consecutive hypo- and oropharyngeal HNC patients, clinically treated with primary IMPT at a Varian ProBeam unit (Varian Medical Systems, Palo Alto, CA) between March 2021 and August 2022, were retrospectively included. Informed consent was obtained from all patients for the retrospective use of their clinical data. Patients were treated with curative intent and irradiated with 7000 cGy (relative biological effectiveness (RBE)=1.1) to the primary Clinical Target Volume (CTV7000), and 5425 cGy (RBE=1.1) to the bilateral elective CTVs (CTV5425) in 35 fractions. See Table 1 for patient and tumor characteristics.

Table 1 Patient and tumor characteristics

		Test patients (n=30)	Tuning patients (n=3)
Age (years)	Median (range)	63 (43-77)	56 (51-67)
Sex (n)	Female	7	-
	Male	23	3
Tumor location (n)	Hypopharynx	2	-
	Oropharynx	28	3
CTV7000 volume (cm^3)	Median (range)	85.4 (16.9-246.6)	76.0 (58.9-126.1)
T-stage (n)	T1	6	1
	T2	16	2
	T3	3	-
	T4*	5	-
N-stage (n)	N0	4	-
	N1	12	1
	N2**	12	2
	N3***	2	-

*T4a, T4b, T4NOS

** N2a, N2b, N2c, N2NOS

***N3a, N3b

Clinical manual planning

The clinically delivered plans were previously manually generated in the clinical treatment planning system (RayStation v10B, RaySearch Laboratories AB, Stockholm, Sweden). A four-field beam configuration (template angles: 50°, 150°, 210°, 310°) including a 5 cm range shifter per beam was used. The anterior fields could be adjusted by the involved planner by 5-10° to avoid irradiation through dental fillings. Spot avoidance volumes were created to prevent spots passing through metal dental fillings. Traversing of shoulders was prevented with avoidance structures comprising the caudal parts of CTVs. CTV doses were robustly optimized for 21 scenarios with a 3 mm isotropic setup uncertainty and ±3% density uncertainty. Doses to OARs were minimized in line with the OAR planning goals in Table 2. The clinical IMPT plans were optimized and calculated using the RayStation Monte Carlo dose engine with 1.0% statistical uncertainty.

Table 2 Clinical treatment planning goals

Target/OAR		Dose volume goal	Dmean
CTV7000		D98%≥95%* D2%≤107%**	
CTV5425		D98%≥95%*	
Parotids			≤2600 cGy
Submandibular glands			≤3500 cGy
Oral cavity			≤2800 cGy
PCMs			≤4000 cGy
Larynx			≤4000 cGy
Cochlea			≤4500 cGy
Brainstem	Core Surface	D0.03cm^3≤5400 cGy** D0.03cm^3≤6000 cGy**	
Spinal cord	Core Surface	D0.03cm^3≤5000 cGy** D0.03cm^3≤6000 cGy**	
Mandible		D2%≤7000 cGy**	

PCMs: Pharyngeal constrictor muscles (delineated separately the superior, medius and superior part)

* Also in the vw-min dose

** Also in the vw-max dose

Automated planning

Wish-list tuning

Configuration of the Erasmus-iCycle wish-list aimed at automated generation of clinically acceptable robust IMPT plans, with at least similar plan quality as the manually generated clinical plans. Definition of an initial wish-list is based on the planning protocol (Table 2) and discussions with clinicians and planners, whereafter this initial wish-list is applied to a limited number of tuning patients. The IMPT plans of three clinical patients, so called tuning patients, were randomly selected from the 33 available IMPT patients plans and used to guide the iterative wish-list tuning process. During the iterative wish-list tuning, plans were evaluated, and if further enhancements are considered feasible, the wish-list is updated. See the appendix of Heijmen et al. [20] for a more detailed description of the process of wish-list tuning. The wish-list used in this study is presented in Table S1 of the supplementary materials.

Automated plan generation

After wish-list tuning, automated plans were generated for the remaining 30 HNC 'test' patients. Automated IMPT plan generation was performed with the recently proposed 'SISS-MCO' method.[28] In SISS-MCO, Sparsity Induced Spot Selection (SISS) precedes final MCO with Erasmus-iCycle. In SISS-MCO, scenario-based minimax robust optimization[33] of CTV doses was applied, using the same 21 scenarios as in clinical manual planning (above). All plans were generated with the same beam arrangement as clinically used. As in clinical manual planning, a 5 cm range shifter was used, and applied contours and avoidance structures were the same as those used in clinical planning. During optimizations, machine parameters of the applied Varian treatment unit were taken into consideration for deliverability, with a minimal monitor unit (MU) per spot constraint of 3 MU and a maximum of 80 MU. Doses were computed with the ASTROID dose engine [34], which was tuned for accurate dose prediction for the applied treatment unit [32].

Plan evaluations and comparisons

For the 30 test patients, automatically generated IMPT plans were compared to the corresponding manually generated clinical IMPT plans. For the CTVs, coverage was evaluated for nominal dose and for voxel-wise minimum (vw-min) dose, constructed from 28 scenarios with 3 mm setup error and ±3% density uncertainty. Both for the nominal plans and the vw-min dose distributions, the clinical D98%≥95% coverage criterion was used, in accordance to the DUPROTON protocol.[35] For CTV7000, the D2%≤107% criterion was applied to nominal dose, and voxel-wise maximum (vw-max) dose distributions, again constructed from the 28 robustness scenarios. Doses in parallel OARs were assessed using mean dose (Dmean) in nominal plans, while for serial OARs, D0.03cm^3 or D2% in nominal and vw-max plans were reported. To evaluate the overall dose to the tissues, the volume of the body that received 5Gy was evaluated (V5Gy). Normal tissue complication probability (NTCP) for xerostomia (grade ≥2 and grade ≥3) and dysphagia (grade ≥2 and grade ≥3) were calculated for nominal plans, using a baseline score of zero for both the clinical and automated plans (Supplementary materials, Table S2).[36] Wilcoxon signed rank tests were performed to assess statistical significance of observed differences between automated and manual planning ($p<0.05$).

RESULTS

Automatically generated IMPT plans with Erasmus-iCycle had similar (robust) target coverage D98%, statistically significant lower doses to OARs and statistically significant lower NTCP for xerostomia and dysphagia as compared to clinical IMPT plans.

Figure 1 shows an example of dose distributions from automated and manual plans, where in the automatically generated plans the isodose lines conform better to the targets, resulting in lower doses in the surrounding healthy tissues, especially in the posterior part of the neck.

For the nominal scenario, the D98%≥95% criterion was fulfilled in all 30 automated and all clinical plans, for both CTVs. For CTV7000, the median D98% of the nominal automated plans was 24 cGy lower than for the nominal clinical plans, while for CTV5425 the automated plans showed a 40 cGy higher median D98% (Table 3). In two clinical plans the D98%≥95% criterion for vw-min dose was not fulfilled for CTV7000, while the criterion was respected in all automated plans for both CTVs. For none of the two CTVs, differences between automated and manual planning in vw-min D98% were statistically significant (Table 3). The D2% was slightly improved in the automated plans as compared to the clinical plans for both nominal and vw-max dose. The median D2% for CTV7000 in the automated plans was 79 cGy lower in the nominal dose and 84 cGy lower for the vw-max dose as compared to the clinical plans. For the nominal dose, the D2%≤107% criterion was fulfilled in all 30 automated and all clinical plans, and for the vw-max dose, all 30 automated plans respected this constraint while it was exceeded in two clinical plans.

For all evaluated OARs, delivered dose was favorable in the automated plans with all $p<0.001$ (Table 3). Observed reductions in median dose parameters compared to manual planning were: 250/282 cGy for contralateral/ipsilateral parotid Dmean, 408/242 cGy for contralateral/ipsilateral submandibular gland Dmean, 115 cGy for oral cavity Dmean, 430/578/667 cGy for pharyngeal constrictor muscle (PCM) superior/medius/inferior Dmean, 494 cGy for larynx Dmean, 224/198 cGy for mandible D2% and vw-min D2%. For the brainstem, dose reductions in the automated plans compared to the manual plans were 1106/1444 cGy for core/surface D0.03cm^3 and 1478/1590 cGy core/surface vw-max D0.03cm^3. For the spinal cord, dose reductions were 1629/1795 cGy for core/surface D0.03cm^3 and 1793/1758 cGy for core/surface vw-max D0.03cm^3. The volume of the body that received 5Gy was reduced by 946 cm^3 in the automated plans compared to the manual plans.

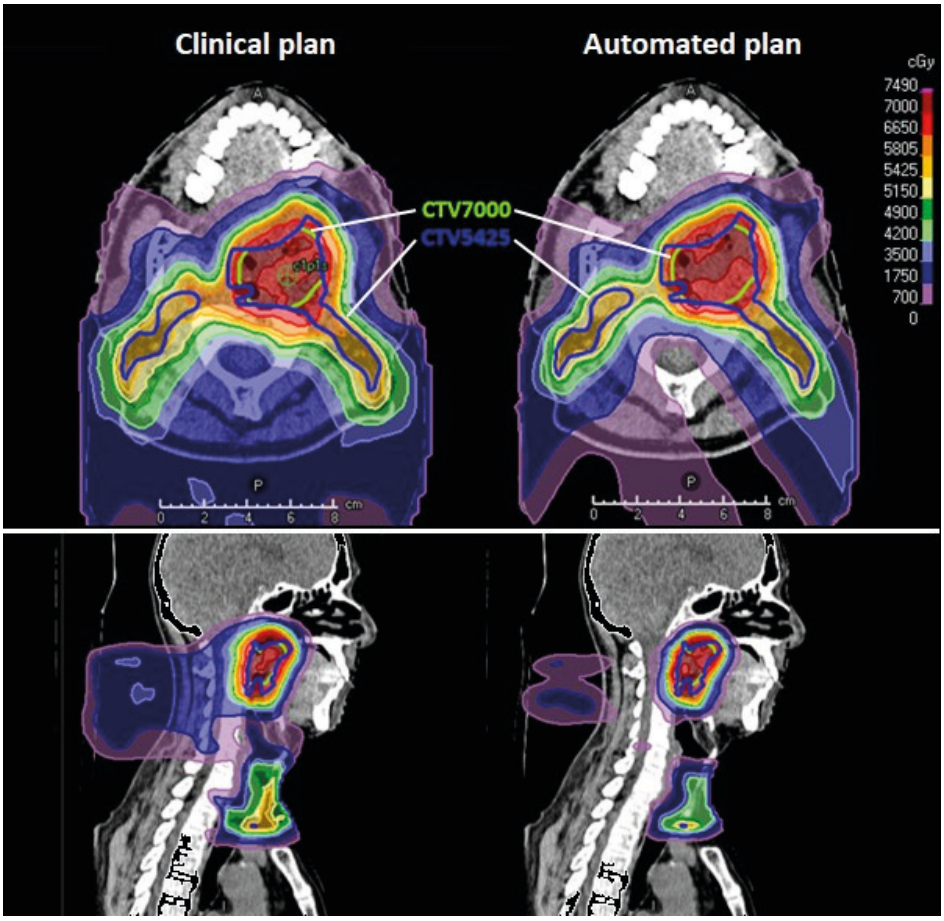

Figure 1 Transversal- (top row) and sagittal cross-sections (bottom row) of a CT with the nominal dose distribution of a clinical plan (left) and an automated plan (right) of one representative patient (patient 30). Note that in the automated plan, the isodose lines conform better to the targets with less dose to the surrounding healthy tissues.

Table 3 Dose/volume metrics (median and range) for the CTVs and OARs, and predicted NTCPs for the clinical and automated plans. In the last column the p-value outcome of the Wilcoxon signed rank tests.

	Metric[+]	Clinical plan median (min-max)	Automated plan median (min-max)	p-value
CTV7000	D98%≥95% (6650 [cGy])	6889 (6821-6933)	6865 (6792-6889)	<0.001*
	vw-min D98%≥95% (6650 [cGy])	6687 (6634-6788)	6707 (6656-6736)	0.131
	D2%≤107% (7490 [cGy])	7305 (7242-7346)	7226 (7049-7259)	<0.001*
	vw-max D2%≤107% (7490 [cGy])	7428 (7355-7501)	7344 (7128-7386)	<0.001*
CTV5425	D98%≥95% (5150 [cGy])	5371 (5342-5427)	5411 (5357-5516)	<0.001*
	vw-min D98%≥95% (5150 [cGy])	5196 (5153-5238)	5193 (5156-5264)	0.572
Parotid contra	Dmean [cGy]	1537 (7-3546)	1287 (0-3093)	<0.001*
Parotid ipsi	Dmean [cGy]	2650 (1430-6266)	2368 (1147-5998)	<0.001*
SMG contra	Dmean [cGy]	3753 (14-5418)	3345 (0-5002)	<0.001*
SMG ipsi	Dmean [cGy]	6287 (4479-6954)	6045 (3806-6867)	<0.001*
Oral cavity	Dmean [cGy]	2793 (597-5144)	2678 (707-4876)	<0.001*
PCM superior	Dmean [cGy]	5663 (3521-7084)	5233 (2975-6936)	<0.001*
PCM medius	Dmean [cGy]	4854 (2637-7053)	4276 (2282-6954)	<0.001*
PCM inferior	Dmean [cGy]	2348 (1178-6678)	1681 (821-6533)	<0.001*
Larynx	Dmean [cGy]	3919 (1674-6725)	3425 (930-6281)	<0.001*
Cochlea left	Dmean [cGy]	145 (13-1995)	7 (0-750)	<0.001*
Cochlea right	Dmean [cGy]	308 (15-2158)	48 (0-866)	<0.001*
Brainstem	Core D0.03cm^3 [cGy]	1201 (125-3054)	95 (0-738)	<0.001*
	vw-max Core D0.03cm^3 [cGy]	1684 (208-3495)	206 (0-841)	<0.001*
	Surface D0.03cm^3 [cGy]	1639 (177-3424)	195 (0-880)	<0.001*
	vw-max Surface D0.03cm^3 [cGy]	2020 (260-3822)	430 (2-1375)	<0.001*
Spinal cord	Core D0.03cm^3 [cGy]	2485 (1024-3588)	856 (349-2014)	<0.001*
	vw-max Core D0.03cm^3 [cGy]	2798 (1139-4084)	1005 (384-2448)	<0.001*
	Surface D0.03cm^3 [cGy]	2741 (1092-4099)	946 (363-2299)	<0.001*
	vw-max Surface D0.03cm^3 [cGy]	3226 (1388-4570)	1468 (587-3076)	<0.001*
Mandible	D2% [cGy]	6687 (4369-6959)	6463 (3302-6887)	<0.001*
	vw-max D2% [cGy]	6913 (4980-7059)	6715 (4043-6996)	<0.001*
Body	V5Gy [cm^3]	3342 (1478-5122)	2396 (1233-3401)	<0.001*
NTCP	Grade ≥2 [%]	38.0 (26.5-50.3)	34.4 (27.7-47.7)	<0.001*
Xerostomia	Grade ≥3 [%]	10.2 (6.7-14.8)	9.0 (6.2-13.7)	<0.001*
NTCP	Grade ≥2 [%]	15.0 (8.8-30.3)	11.8 (6.9-26.7)	<0.001*
Dysphagia	Grade ≥3 [%]	2.8 (1.4-14.2)	1.8 (0.9-12.3)	<0.001*

[+]Metrics are provided for nominal plans, unless stated otherwise by 'vw-min' or 'vw-max'.

SMG=Submandibular gland, PCM=Pharyngeal constrictor muscle

Figure 2 shows per patient, the differences between automated and manual plans in the OAR Dmean that were used in the NTCP calculations. For patients 1 and 23, the advantage of automated planning became maybe less clear, while for all other 28 patients, automated planning resulted in clearly favorable OAR Dmean. The achieved reductions in OAR Dmean with automated planning led to statistically significant ($p<0.001$) decreases in the NTCP for xerostomia grade ≥2 and grade ≥3 and dysphagia grade ≥2 and grade ≥3 (Table 3, two last rows and Figure 3).

The automatically generated IMPT plans by Erasmus-iCycle were generated within several hours, similar to clinical manual planning times of approximately 6 hours.

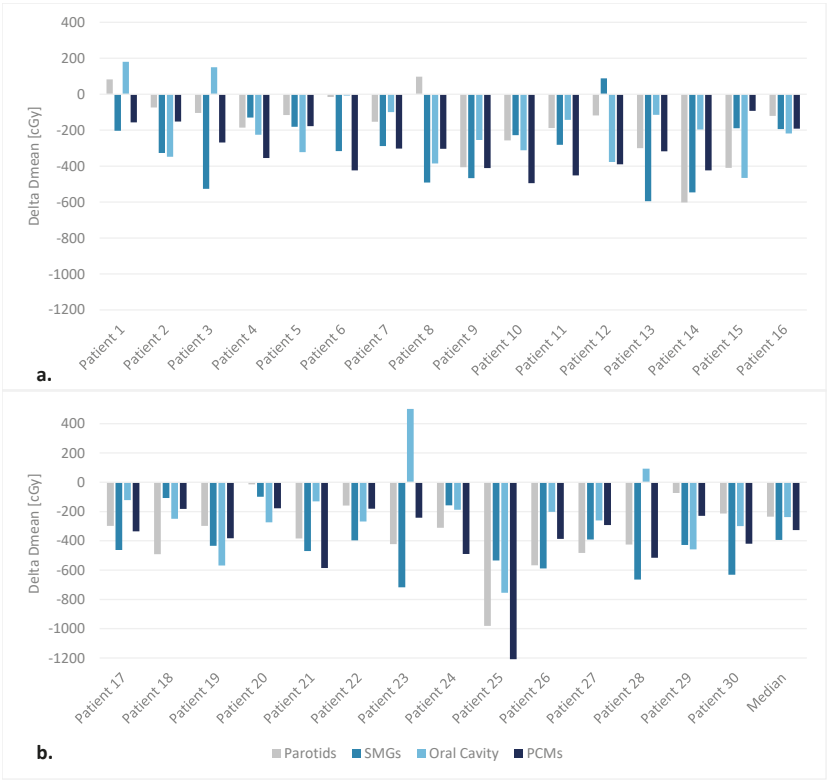

Figure 2 Per patient, differences between the automated and clinical plan in OAR Dmean. Parotids=mean of both parotid glands, SMGs=mean of both submandibular glands, PCMs=mean of the whole pharyngeal constrictor muscle. **a.)** Patient 1 till 16, **b.)** Patient 17 till 30 and population median differences. Positive values indicate lower mean doses in the clinical plans, negative values indicate lower mean doses in automated plans. For patient 23, a dose increase in the oral cavity was seen, with higher total dose reductions in parotids, SMGs and PCMs.

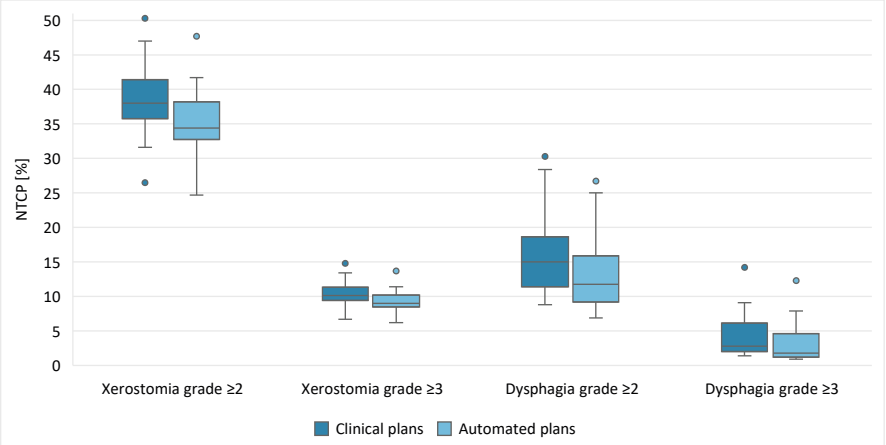

Figure 3 Boxplots of the xerostomia and dysphagia NTCP for the clinical and automated plans. The boxes represent the 25th to 75th percentile of the data, the median is shown with the horizontal line and the whiskers extend to the 10th and 90th percentiles.

DISCUSSION

In this study, robust IMPT plans, generated with fully automated multi-criteria optimization as implemented in Erasmus-iCycle, were compared to manually created, clinically applied robust IMPT plans for HNC patients. Erasmus-iCycle was able to automatically produce clinically acceptable robust IMPT plans for all 30 test patients with comparable (robust) target coverage as the clinical plans, but highly favorable OAR dose parameters (all median differences $p<0.001$). The mean dose (Dmean) reductions to the parallel OARs resulted in statistically significant NTCP reductions for xerostomia and dysphagia compared to clinical plans.

Whereas in all 30 automated plans the clinical robust target D98% goals were achieved, this was not the case for two of the clinical plans. For 28/30 patients, the automatically generated plan was clearly favorable for the obtained Dmean to OARs. For the other two patients, there were gains and losses in dose to OARs with automated planning (Fig. 2). The largest median dose reductions were up to 667 cGy in the Dmean to the inferior PCM. Large dose reductions were also observed in serial OARs, e.g. a median reduction in D0.03cm^3 of the spinal cord surface of 1795 cGy. Also the posterior part of the neck was much better spared with automated planning.

To the best of our knowledge, this is the first validation study for fully automated MCO in IMPT, and it shows clinically relevant plan quality enhancements compared to manual planning. Advantages of automated MCO compared to manual planning

have been observed previously in external photon beam radiotherapy.[17–25] In line with our results, automated photon plans for HNC were superior to the manual plans in 97% of the cases.[37] Apparently, it is difficult for human planners to consistently reach the quality obtained by the wish-list driven systematic plan generation in automated MCO. There are several factors that can contribute to this, including restrictions in available planning times for each individual patient's plan in manual planning, and variations in the human planner experience and planning skills.

With Erasmus-iCycle automated MCO, generated plans are guaranteed Pareto-optimal [12] and all objectives are minimized to the maximum extent, while never violating constraints and taking into account objective priorities (see Table S1). For spinal cord and brainstem, the applied wish-list has maximum dose constraints. In clinical manual planning, a plan is generally accepted if such constraints are met. In this study, we have added low priority objectives (priority 12) to the wish-list for these structures. As a result, the optimizer will at the end of the plan generation reduce doses in these OARs if possible, without deteriorating the higher priority objectives. In this way, not only the constraints are met, but doses are also maximally reduced. Application of the low priority objectives for the brainstem and spinal cord have contributed to the large dose reductions in these structures compared to manual planning. It probably also contributed to the obtained low-neck doses. Priorities 13-15 aimed at conformality improvements without deteriorating any of the higher priority objectives.

The large reductions in observed D0.03cm^3 for the brainstem and spinal cord can be of importance in case of future reirradiations are necessary. In an eventual re-irradiation plan, some overlapping dose in these structures will be possible since the earlier delivered doses were far below the clinical constraints. For the purpose of plan adaptation, Erasmus-iCycle is highly suitable for offline replanning since it is able to automatically produce an adjusted plan with high and consistent plan quality within several hours and without manual intervention. Speeding up this optimization time is part of further investigations.

The obtained NTCP reductions can be of clinical relevance, not only because patients may develop less side effects, but also in case patient selection for IMPT is according to a photon-proton model-based plan comparison.[36] Since NTCP reductions obtained with Erasmus-iCycle potentially lead to larger differences in NTCP between the photon-proton plans, this could increase the chance that patients are eligible for proton therapy. However, the fairest comparison would then be an automated Erasmus-iCycle photon to proton plan comparison. Automation of

photon-proton plan comparisons could speed up the process and prevent delays in starting with the (proton) treatment. An extra added value of fully automated planning are cost reductions because of reduced planning workload, and it can also solve problems in case of scarcity of well-trained planners. For photon therapy, several studies have reported on workload reductions with automated planning. Buschmann et al. observed workload reductions of >70 min compared to manual planning for prostate cancer VMAT.[21] Fjellanger et al. observed reductions in hands-on planning times for lung cancer VMAT from 2-4 h with manual planning to less than 10 min for automated planning.[38] Marrazzo et al. reported a reduction in hands-on planning time for partial breast irradiation from 63 min to 10 min. [39] In any case, we believe that development of planning tools that allow proton centers to maximally exploit the benefit of proton beams is of crucial importance. It is of great interest for the patient and also for society, as scarce and expensive equipment will be used in the best way possible.

Whereas automated treatment planning methods for photons has been increasingly investigated over the past few years as described by Hussein et al. [8], automated planning studies on IMPT are relatively scarce, but a few have been performed previously.[40–43] Delaney et al.[40,41] investigated a commercially available knowledge-based planning method for IMPT HNC patients, and also performed a multi-center study. In line with our results, they found comparable target coverage and reduced OAR dose in the knowledge-based planning plans as compared to the clinical IMPT plans. Van Bruggen et al.[42] also investigated a commercially available method for automated IMPT planning in HNC patients, based on machine learning. They found comparable target coverage and NTCP, and significant higher dose to some OARs in the automated plans as compared to the clinical plans.

Both studies are based on an atlas of earlier treated patients, and the resulting IMPT plans are dependent on the quality of this atlas patients. This contrasts to our Erasmus-iCycle optimization method, which is an a-priori MCO method and able to produce high quality plans which are independent of the quality of earlier applied plans. Taasti et al.[43] investigated an in-house developed fully automated treatment planning method in protons for HNC cases, and extended this with an automated beam angle selection method. Their optimization method is, similar as Erasmus-iCycle, based on hierarchical optimization of constraints. Plans with automated selected beam configuration had reduced dose to the mandible and unspecified healthy tissues as compared to the configuration chosen by the planner.

Although in this paper the beam arrangement in the clinical and Erasmus-iCycle plans were generated with the same four-field beam configuration, it is expected

that the plan quality of IMPT plans can be further improved with the application of automated beam angle selection.[43] This has been shown in various treatment sites for Erasmus-iCycle in photon plans.[12,44–46] Automated beam angle selection will be part of further development for Erasmus-iCycle IMPT.

A possible limitation of our current study is that our clinical plans are calculated using Monte Carlo dose engine, while we use a pencil beam dose calculation algorithm calibrated to the measured beam data from HollandPTC in our automated plans. Ideally, exactly the same dose calculation algorithm should have been used. However, in literature we found that differences between analytical dose calculation and Monte Carlo in IMPT for HNC are small. Yepes et et al.[47] investigated in a large HNC cohort (n=125) the differences between analytical and Monte Carlo dose calculations. They found that target coverage was predicted slightly higher with analytical computations as compared to Monte Carlo computations, and for all OARs, median differences in dose computations were within 1%.

Erasmus-iCycle IMPT plans respect the clinically applied minimum and maximum allowed MU per spot, minimum and maximum available proton energies and it adheres to the clinically applied separation between energy layers. However, currently the system is not coupled to our treatment unit. Clinical application requires regulatory procedures following the European Union Medical Device Regulation (EU MDR) (https://eumdr.com/) that aims at guaranteeing safe usage of medical devices. The rules apply both for in-house developed and commercial software/devices.

This study is the first validation of the Erasmus-iCycle for clinical IMPT planning. Erasmus-iCycle IMPT plans showed excellent performance for HNC regarding dosimetric plan parameters and predicted NTCPs. Further validation studies for other tumor sites should be done.

CONCLUSIONS

For HNC patients, fully automated MCO resulted in clinically acceptable robust IMPT plans with comparable target coverage as the clinical, manually generated plans, but with highly favorable OAR dose parameters. Achieved OAR Dmean reductions resulted in statistically significant reductions in predicted xerostomia and dysphagia NTCPs. For the vast majority of patients, the automatically generated plan was clearly favorable for obtained OAR Dmean. Large dose reductions were also observed in serial OARs and in the posterior part of the neck.

REFERENCES

1. Parvathaneni U, Laramore GE, Liao JJ. Technical Advances and Pitfalls in Head and Neck Radiotherapy. J Oncol 2012;2012:13. https://doi.org/10.1155/2012/597467.
2. Blanchard P, Gunn GB, Lin A, Foote RL, Lee NY, Frank SJ. Proton Therapy for Head and Neck Cancers. Semin Radiat Oncol 2018;28:53–63. https://doi.org/10.1016/J.SEMRADONC.2017.08.004.
3. Water TA van de, Bijl HP, Schilstra C, Pijls-Johannesma M, Langendijk JA. The Potential Benefit of Radiotherapy with Protons in Head and Neck Cancer with Respect to Normal Tissue Sparing: A Systematic Review of Literature. Oncologist 2011;16:366. https://doi.org/10.1634/THEONCOLOGIST.2010-0171.
4. Liu W, Frank SJ, Li X, Li Y, Park PC, Dong L, et al. Effectiveness of robust optimization in intensity-modulated proton therapy planning for head and neck cancers. Med Phys 2013;40. https://doi.org/10.1118/1.4801899.
5. Unkelbach J, Paganetti H. Robust Proton Treatment Planning: Physical and Biological Optimization. Semin Radiat Oncol 2018;28:88–96. https://doi.org/10.1016/J.SEMRADONC.2017.11.005.
6. 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. https://doi.org/10.1016/J.PRRO.2016.04.005.
7. 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.
8. Hussein M, Heijmen BJM, Verellen D, Nisbet A. Automation in intensity modulated radiotherapy treatment planning-a review of recent innovations. Br J Radiol 2018;91. https://doi.org/10.1259/BJR.20180270/ASSET/IMAGES/LARGE/BJR.20180270.G004.JPEG.
9. Fogliata A, Belosi F, Clivio A, Navarria P, Nicolini G, Scorsetti M, 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. https://doi.org/10.1016/J.RADONC.2014.11.009.
10. McIntosh C, Welch M, McNiven A, Jaffray DA, Purdie TG. Fully automated treatment planning for head and neck radiotherapy using a voxel-based dose prediction and dose mimicking method. Phys Med Biol 2017;62:5926. https://doi.org/10.1088/1361-6560/AA71F8.
11. Monz M, Küfer KH, Bortfeld TR, Thieke C. Pareto navigation—algorithmic foundation of interactive multi-criteria IMRT planning. Phys Med Biol 2008;53:985. https://doi.org/10.1088/0031-9155/53/4/011.
12. Breedveld S, Storchi PRM, Voet PWJ, Heijmen BJM. iCycle: Integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar IMRT plans. Med Phys 2012;39:951–63. https://doi.org/10.1118/1.3676689.
13. Breedveld S, Storchi PRM, Heijmen BJM. The equivalence of multi-criteria methods for radiotherapy plan optimization. Phys Med Biol 2009;54:7199–209. https://doi.org/10.1088/0031-9155/54/23/011.

14. Hansen CR, Bertelsen A, Hazell I, Zukauskaite R, Gyldenkerne N, Johansen J, et al. Automatic treatment planning improves the clinical quality of head and neck cancer treatment plans. Clin Transl Radiat Oncol 2016;1:2–8. https://doi.org/10.1016/J.CTRO.2016.08.001.

15. Krayenbuehl J, Norton I, Studer G, Guckenberger M. Evaluation of an automated knowledge based treatment planning system for head and neck. Radiat Oncol 2015;10:1–8. https://doi.org/10.1186/S13014-015-0533-2/FIGURES/3.

16. Hazell I, Bzdusek K, Kumar P, Hansen CR, Bertelsen A, Eriksen JG, et al. Automatic planning of head and neck treatment plans. J Appl Clin Med Phys 2016;17:272–82. https://doi.org/10.1120/JACMP.V17I1.5901.

17. Buergy D, Sharfo AWM, Heijmen BJM, Voet PWJ, Breedveld S, Wenz F, et al. Fully automated treatment planning of spinal metastases - A comparison to manual planning of Volumetric Modulated Arc Therapy for conventionally fractionated irradiation. Radiat Oncol 2017;12. https://doi.org/10.1186/S13014-017-0767-2.

18. Sharfo AWM, Breedveld S, Voet PWJ, Heijkoop ST, Mens JWM, Hoogeman MS, et al. Validation of Fully Automated VMAT Plan Generation for Library-Based Plan-of-the-Day Cervical Cancer Radiotherapy. PLoS One 2016;11. https://doi.org/10.1371/JOURNAL.PONE.0169202.

19. 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 2014;88:1175–9. https://doi.org/10.1016/J.IJROBP.2013.12.046.

20. Heijmen B, Voet P, Fransen D, Penninkhof J, Milder M, Akhiat H, et al. Fully automated, multi-criterial planning for Volumetric Modulated Arc Therapy – An international multi-center validation for prostate cancer. Radiother Oncol 2018;128:343–8. https://doi.org/10.1016/J.RADONC.2018.06.023.

21. Buschmann M, Sharfo AWM, Penninkhof J, Seppenwoolde Y, Goldner G, Georg D, et al. Automated volumetric modulated arc therapy planning for whole pelvic prostate radiotherapy. Strahlenther Onkol 2018;194:333–42. https://doi.org/10.1007/S00066-017-1246-2.

22. Sharfo AWM, Dirkx MLP, Bijman RG, Schillemans W, Breedveld S, Aluwini S, et al. Late toxicity in the randomized multicenter HYPRO trial for prostate cancer analyzed with automated treatment planning. Radiother Oncol 2018;128:349–56. https://doi.org/10.1016/J.RADONC.2018.05.028.

23. Sharfo AWM, Stieler F, Kupfer O, Heijmen BJM, Dirkx MLP, Breedveld S, et al. Automated VMAT planning for postoperative adjuvant treatment of advanced gastric cancer. Radiat Oncol 2018;13. https://doi.org/10.1186/S13014-018-1032-Z.

24. Bijman R, Rossi L, Janssen T, de Ruiter P, Carbaat C, van Triest B, et al. First system for fully-automated multi-criterial treatment planning for a high-magnetic field MR-Linac applied to rectal cancer. Acta Oncol 2020;59:926–32. https://doi.org/10.1080/0284186X.2020.1766697.

25. Bijman R, Sharfo AW, Rossi L, Breedveld S, Heijmen B. Pre-clinical validation of a novel system for fully-automated treatment planning. Radiother Oncol 2021;158:253–61. https://doi.org/10.1016/J.RADONC.2021.03.003.

26. Van De Water S, Kraan AC, Breedveld S, Schillemans W, Teguh DN, Kooy HM, et al. Improved efficiency of multi-criteria IMPT treatment planning using iterative resampling of randomly placed pencil beams. Phys Med Biol 2013;58:6969. https://doi.org/10.1088/0031-9155/58/19/6969.

27. Van De Water S, Kooy HM, Heijmen BJM, Hoogeman MS. Shortening Delivery Times of Intensity Modulated Proton Therapy by Reducing Proton Energy Layers During Treatment Plan Optimization. Int J Radiat Oncol 2015;92:460–8. https://doi.org/10.1016/J.IJROBP.2015.01.031.

28. Kong W, Oud M, Habraken SJM, Huiskes M, Astreinidou E, Rasch CRN, et al. OC-0287 Fast and fully-automated robust multi-criterial IMPT planning via sparsity induced spot selection. Radiother. Oncol., 2023, p. S222–3.

29. van de Water S, van Dam I, Schaart DR, Al-Mamgani A, Heijmen BJM, Hoogeman MS. The price of robustness; impact of worst-case optimization on organ-at-risk dose and complication probability in intensity-modulated proton therapy for oropharyngeal cancer patients. Radiother Oncol 2016;120:56–62. https://doi.org/10.1016/J.RADONC.2016.04.038.

30. Van De Water S, Albertini F, Weber DC, Heijmen BJM, Hoogeman MS, Lomax AJ. Anatomical robust optimization to account for nasal cavity filling variation during intensity-modulated proton therapy: a comparison with conventional and adaptive planning strategies. Phys Med Biol 2018;63:025020. https://doi.org/10.1088/1361-6560/AA9C1C.

31. Oud M, Breedveld S, Giżyńska M, Kroesen M, Hutschemaekers S, Habraken S, et al. An online adaptive plan library approach for intensity modulated proton therapy for head and neck cancer. Radiother Oncol 2022;176:68–75. https://doi.org/10.1016/J.RADONC.2022.09.011.

32. Kouwenberg J, Penninkhof J, Habraken S, Zindler J, Hoogeman M, Heijmen B. Model based patient pre-selection for intensity-modulated proton therapy (IMPT) using automated treatment planning and machine learning. Radiother Oncol 2021;158:224–9. https://doi.org/10.1016/J.RADONC.2021.02.034.

33. Fredriksson A, Forsgren A, Hårdemark B. Minimax optimization for handling range and setup uncertainties in proton therapy. Med Phys 2011;38:1672–84. https://doi.org/10.1118/1.3556559.

34. Kooy HM, Clasie BM, Lu HM, Madden TM, Bentefour H, Depauw N, et al. A Case Study in Proton Pencil-Beam Scanning Delivery. Int J Radiat Oncol 2010;76:624–30. https://doi.org/10.1016/J.IJROBP.2009.06.065.

35. Korevaar EW, Habraken SJM, Scandurra D, Kierkels RGJ, Unipan M, Eenink MGC, et al. Practical robustness evaluation in radiotherapy – A photon and proton-proof alternative to PTV-based plan evaluation. Radiother Oncol 2019;141:267–74. https://doi.org/10.1016/J.RADONC.2019.08.005.

36. Langendijk JA, Hoogeman MS, Monshouwer R VM. Landelijk Indicatie Protocol Protonentherapie (versie 2.2) (LIPPv2.2) HOOFD-HALSTUMOREN 2019. https://nvro.nl/publicaties/rapporten (accessed February 24, 2023).

37. Voet PWJ, Dirkx MLP, Breedveld S, Fransen D, Levendag PC, Heijmen BJM. Toward fully automated multicriterial plan generation: A prospective clinical study. Int J Radiat Oncol Biol Phys 2013;85:866–72. https://doi.org/10.1016/J.IJROBP.2012.04.015.

38. Fjellanger K, Hordnes M, Sandvik IM, Sulen TH, Heijmen BJM, Breedveld S, et al. Improving knowledge-based treatment planning for lung cancer radiotherapy with automatic multi-criteria optimized training plans. Acta Oncol 2023. https://doi.org/10.1080/0284186X.2023.2238882.

39. Marrazzo L, Meattini I, Arilli C, Calusi S, Casati M, Talamonti C, et al. Auto-planning for VMAT accelerated partial breast irradiation. Radiother Oncol 2019;132:85–92. https://doi.org/10.1016/J.RADONC.2018.11.006.

40. Delaney AR, Verbakel WF, Lindberg J, Koponen TK, Slotman BJ, Dahele M. Evaluation of an Automated Proton Planning Solution. Cureus 2018;10. https://doi.org/10.7759/CUREUS.3696.

41. Delaney AR, Dong L, Mascia A, Zou W, Zhang Y, Yin L, et al. Automated Knowledge-Based Intensity-Modulated Proton Planning: An International Multicenter Benchmarking Study. Cancers (Basel) 2018;10. https://doi.org/10.3390/CANCERS10110420.

42. van Bruggen IG, Huiskes M, de Vette SPM, Holmström M, Langendijk JA, Both S, et al. Automated Robust Planning for IMPT in Oropharyngeal Cancer Patients Using Machine Learning. Int J Radiat Oncol 2022. https://doi.org/10.1016/J.IJROBP.2022.12.004.

43. Taasti VT, Hong L, Shim JS, Deasy JO, Zarepisheh M. Automating proton treatment planning with beam angle selection using Bayesian optimization. Med Phys 2020;47:3286. https://doi.org/10.1002/MP.14215.

44. Voet PWJ, Breedveld S, Dirkx MLP, Levendag PC, Heijmen BJM. Integrated multicriterial optimization of beam angles and intensity profiles for coplanar and noncoplanar head and neck IMRT and implications for VMAT. Med Phys 2012;39:4858–65. https://doi.org/10.1118/1.4736803.

45. Sharfo AWM, Dirkx MLP, Breedveld S, Romero AM, Heijmen BJM. VMAT plus a few computer-optimized non-coplanar IMRT beams (VMAT+) tested for liver SBRT. Radiother Oncol 2017;123:49–56. https://doi.org/10.1016/J.RADONC.2017.02.018.

46. Rossi L, Sharfo AW, Aluwini S, Dirkx M, Breedveld S, Heijmen B. First fully automated planning solution for robotic radiosurgery – comparison with automatically planned volumetric arc therapy for prostate cancer. Https://DoiOrg/101080/0284186X20181479068 2018;57:1490–8. https://doi.org/10.1080/0284186X.2018.1479068.

47. Yepes P, Adair A, Grosshans D, Mirkovic D, Poenisch F, Titt U, et al. Comparison of Monte Carlo and analytical dose computations for intensity modulated proton therapy. Phys Med Biol 2018;63. https://doi.org/10.1088/1361-6560/AAA845.

SUPPLEMENTARY MATERIAL

Table S1 The Erasmus-iCycle wish-list used in this study

Constraints					
	Structure	Constraint function	Type	Constraint	Robust
	CTV7000	Minimum Dose	Linear	6755 cGy	Yes
	CTV5425	Minimum Dose	Linear	6235 cGy	Yes
	Brainstem	Maximum Dose	Linear	4500 cGy	
	Spinal Cord	Maximum Dose	Linear	4500 cGy	
	MU per spot	Maximum Dose	Linear	80	
Objectives					
Priority	Structure	Aim & objective function	Type	Goal value (Sufficient[#])	Robust
1	CTV7000	↓ Maximum Dose	Linear	7420 cGy	Yes
1	CTV5425	↓ Maximum Dose	Linear	5750 cGy	Yes
2	0-10 mm ring around CTV_combined	↓ Maximum Dose	Linear	5750 cGy	
2	10-15 mm ring around CTV_combined	↓ Maximum Dose	Linear	4880 cGy	
3	Mandible	↓ Maximum Dose	Linear	6900 cGy	
4	Cochlea	↓ Maximum Dose	Mean	4500 Gy (500 cGy)	
5	Brain	↓ Maximum Dose	Linear	100 cGy	
6	Parotids	↓ Maximum Dose	Mean	2600 cGy (100 cGy)	
7	Submandibular glands	↓ Maximum Dose	Mean	3500 cGy (100 cGy)	
8	PCMs	↓ Maximum Dose	Mean	4000 cGy (100 cGy)	
9	Oral cavity	↓ Maximum Dose	Mean	2800 cGy (100 cGy)	
10	Larynx	↓ Maximum Dose	Mean	4000 cGy (100 cGy)	
11	Esophagus	↓ Maximum Dose	Mean	4000 cGy (100 cGy)	
12	Brainstem	↓ Maximum Dose	Linear	4500 cGy (100 cGy)	
12	Spinal cord	↓ Maximum Dose	Linear	4500 cGy (100 cGy)	
13	Mandible	↓ Maximum Dose	Linear	6900 cGy (100 cGy)	
14	0-10 mm ring around CTV_combined	↓ Maximum Dose	Mean	100 cGy	
14	10-15 mm ring around CTV_combined	↓ Maximum Dose	Mean	100 cGy	
14	15-25 mm ring around CTV_combined	↓ Maximum Dose	Mean	100 cGy	

[#] = If no value; sufficient value is the goal value, ↓ = minimization, ↑ = maximization, CTV_combined = CTV5425 combined with CTV7000+10mm, PCMs: Pharyngeal constrictor muscles (delineated separately the superior, medius and inferior part)

Table S2 Regression coefficients for primary tumors for calculations of the normal tissue complication probability (NTCP) according to the Dutch National Indication Protocol for Proton Therapy.[1]

Variables	Endpoint (6 Months after Radiotherapy)			
	Xerostomia Grade ≥2	Dysphagia Grade ≥3	Xerostomia Grade ≥2	Dysphagia Grade ≥3
Constant (B0)	−2.2951	−3.7286	−4.0536	−7.6174
Slopes				
√Dmean Parotid ipsilateral + √Dmean Parotid contralateral	0.0996	0.0855		
Dmean submandibular bilateral	0.0182	0.0156		
Dmean Oral cavity			0.0300	0.0259
Dmean PCM superior			0.0236	0.0203
Dmean PCM medius			0.0095	0.0303
Dmean PCM inferior			0.0133	0.0341
Baseline score				
Baseline xerostomia: None (EORTC QLQ-H&N35—Q41: score 1)	0.0000	0.0000		
Baseline xerostomia: A little (EORTC QLQ-H&N35—Q41: score 2)	0.4950	0.4249		
Baseline xerostomia: Quite (EORTC QLQ-H&N35—Q41: score 3–4)	1.2070	1.0361		
Baseline grade 0–1 dysphagia (normal foods)			0.0000	0.0000
Baseline grade 2 dysphagia (soft foods)			0.9382	0.5738
Baseline grade 3–4 dysphagia (liquid foods or TFD)			1.2900	1.4718
Tumor location				
Oral Cavity			0.0000	0.0000
Pharynx			−0.6281	0.0387
Larynx			−0.7711	−0.5303

PCM: Pharyngeal constrictor muscle

The NTCP can be calculated with this formula:

$$NTCP = \left(1 + e^{-(Constant + \sum Slope \cdot D_{mean} + Baseline\ score + Tumor\ location)}\right)^{-1}$$

1. Langendijk JA, Hoogeman MS, Monshouwer R VM. Landelijk Indicatie Protocol Protonentherapie (versie 2.2) (LIPPv2.2) HOOFD-HALSTUMOREN 2019. https://nvro.nl/publicaties/rapporten (accessed February 24, 2023).