

University of Groningen

Development of advanced preselection tools to reduce redundant plan comparisons in model-based selection of head and neck cancer patients for proton therapy

Tambas, Makhbule; der Laan, Hans P van; Rutgers, Wouter; van den Hoek, Johanna G M; Oldehinkel, Edwin; Meijer, Tineke W H; van der Schaaf, Arjen; Scandurra, Daniel; Free, Jeffrey; Both, Stefan

Published in:
Radiotherapy and Oncology

DOI:
[10.1016/j.radonc.2021.04.012](https://doi.org/10.1016/j.radonc.2021.04.012)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Tambas, M., der Laan, H. P. V., Rutgers, W., van den Hoek, J. G. M., Oldehinkel, E., Meijer, T. W. H., van der Schaaf, A., Scandurra, D., Free, J., Both, S., Steenbakkens, R. J. H. M., & Langendijk, J. A. (2021). Development of advanced preselection tools to reduce redundant plan comparisons in model-based selection of head and neck cancer patients for proton therapy. *Radiotherapy and Oncology*, 160, 61-68. <https://doi.org/10.1016/j.radonc.2021.04.012>

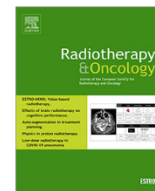
Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Original Article

Development of advanced preselection tools to reduce redundant plan comparisons in model-based selection of head and neck cancer patients for proton therapy



Makbule Tambas*, Hans P. van der Laan, Wouter Rutgers, Johanna G.M. van den Hoek, Edwin Oldehinkel, Tineke W.H. Meijer, Arjen van der Schaaf, Daniel Scandurra, Jeffrey Free, Stefan Both, Roel J.H.M. Steenbakkers, Johannes A. Langendijk

University of Groningen, University Medical Center Groningen, Department of Radiation Oncology, Groningen, the Netherlands

ARTICLE INFO

Article history:

Received 9 December 2020

Received in revised form 6 April 2021

Accepted 9 April 2021

Available online 21 April 2021

Keywords:

Proton therapy

Head and neck cancer

Patient selection

Preselection

IMPT

Plan comparison

ABSTRACT

Purpose: In the Netherlands, head and neck cancer (HNC) patients are selected for proton therapy (PT) based on estimated normal tissue complication probability differences (Δ NTCP) between photons and protons, which requires a plan comparison (VMAT vs. IMPT). We aimed to develop tools to improve patient selection for plan comparisons.

Methods: This prospective study consisted of 141 consecutive patients in which a plan comparison was done. IMPT plans of patients not qualifying for PT were classified as 'redundant'. To prevent redundant IMPT planning, 5 methods that were primarily based on regression models were developed to predict IMPT D_{mean} to OARs, by using data from VMAT plans and volumetric data from delineated targets and OARs. Then, actual and predicted plan comparison outcomes were compared. The endpoint was being selected for proton therapy.

Results: Seventy out of 141 patients (49.6%) qualified for PT. Using the developed preselection tools, redundant IMPT planning could have been prevented in 49–68% of the remaining 71 patients not qualifying for PT (=specificity) when the sensitivity of all methods was fixed to 100%, i.e., no false negative cases (positive predictive value range: 57–68%, negative predictive value: 100%).

Conclusion: The advanced preselection tools, which uses volume and VMAT dose data, prevented labour intensive creation of IMPT plans in up to 68% of non-qualifying patients for PT. No patients qualifying for PT would have been incorrectly denied a plan comparison. This method contributes significantly to a more cost-effective model-based selection of HNC patients for PT.

© 2021 The Author(s). Published by Elsevier B.V. Radiotherapy and Oncology 160 (2021) 61–68 This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

In the Netherlands, patients with head and neck cancer (HNC) are selected for proton therapy using a model-based approach. In each patient, a photon vs. proton (Volumetric-Modulated Arc Therapy (VMAT) vs. Intensity-Modulated Proton Therapy (IMPT)) plan comparison is made to determine the expected normal tissue complication probability difference (Δ NTCP) between the two modalities using normal tissue complication probability (NTCP) models

Abbreviations: HNC, head and neck cancer; VMAT, volumetric-modulated arc therapy; IMPT, intensity modulated proton therapy; OARs, organ at risks; NTCP, normal tissue complication probability; Δ NTCP, NTCP difference; NIPP, National Indication Protocol Proton therapy; Σ Δ NTCP, the summed risk reduction; TPS, treatment planning system.

* Corresponding author at: University of Groningen, University Medical Center Groningen, Department of Radiation Oncology, Hanzeplein 1, 9713 GZ Groningen, the Netherlands.

E-mail address: m.tambas@umcg.nl (M. Tambas).

<https://doi.org/10.1016/j.radonc.2021.04.012>

0167-8140/© 2021 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

[1–3]. According to the Dutch National Indication Protocol for Proton Therapy (NIPP), three NTCP models (\geq Grade 2 xerostomia [4], \geq Grade 2 dysphagia [5] and tube feeding dependence [6]) and three Δ NTCP thresholds are used for patient selection: 1) Δ NTCP \geq 10% for Grade \geq 2; 2) Δ NTCP \geq 5% for Grade \geq 3; and 3) Summed risk reduction (Σ Δ NTCP) \geq 15% for the two Grade \geq 2 toxicities. Patients who meet any of these Δ NTCP thresholds qualify for proton therapy and treatment costs are completely reimbursed by insurance companies [3,7].

Model-based selection identifies patients that are expected to benefit most from proton therapy. However, the main disadvantage is that for each individual patient, an in-silico treatment planning comparison is required, which is a labour-intensive and time-consuming process. In the Netherlands, the expected rate of patients that will qualify for proton therapy in HNC is around 30–40% [3]. Consequently, in 60–70% of the HNC cases, the proton

plan will not be used for actual treatment and can be considered redundant for the clinical workload.

In the current workflow, patients are selected for a plan comparison by means of a basic preselection tool which uses the dose parameters of the clinical VMAT plan as input. The ΔNTCP thresholds are calculated assuming all OARs receive zero dose when protons are used. This workflow results in a plan comparison for many patients, even for those who are unlikely to qualify for proton therapy [7].

Given the complexity of IMPT planning, it is more resource demanding than VMAT planning [8]. In our clinic, an IMPT plan takes approximately 2.5 days from initiation of planning to the review and approval by the multidisciplinary team while this process takes approximately half a day for VMAT. Treatment plan comparison, when redundant, may lead to a delay in the start of treatment, which may in turn negatively affect survival outcome [9–11]. Therefore, reducing redundant planning comparisons is also in the patient's interest.

In order to accelerate the model-based selection procedure, to increase cost effectiveness and to lower clinical workload, a more advanced preselection tool is needed. Therefore, we aimed to investigate various methods to predict a realistic IMPT OAR dose profile and develop a preselection tool that is straightforward, easy to implement and more accurately identifies patients for a plan comparison with the highest possible probability to qualify for proton therapy.

Materials and methods

Patients

In this study, 172 consecutive HNC patients who went through model-based selection were included. All data were collected prospectively. The clinical characteristics of the patients are shown in [Appendix A: Table A1 in the supplementary data](#). The majority of the population consisted of patients with locally advanced (Stage III–IV; 82%) disease located in the pharyngeal (58%) and laryngeal (38%) region.

Patients were treated with definitive VMAT 70 Gy and 54.25 Gy \pm systemic treatment or IMPT 70 Gy_{RBE} and 54.25 Gy_{RBE} \pm systemic treatment for high risk (PTV_7000) and prophylactic target volumes (PTV_5425) in 35 fractions with a simultaneous integrated boost technique. The same target dose was prescribed for both VMAT and IMPT plans. The VMAT plan was optimized such that at least 98% of the Planning Target Volume (PTV) was covered by 95% of the prescribed dose, while dose to the 2% of PTV was kept below 107%. The IMPT plans were created using pencil beam scanning (PBS) with a minimax robust optimization procedure (range uncertainty: $\pm 3.0\%$, set-up uncertainty: 3 mm, robustness evaluation: 28 error scenarios (including 14 translations and $\pm 3\%$ range uncertainties). Plans were evaluated using the voxel-wise minimum for CTV-coverage) in RayStation (v6.1 and v8B, RaySearch Laboratories AB, Stockholm, Sweden) [12,13]. plan robustness was evaluated for 28 error scenarios (including 14 translations and $\pm 3\%$ range uncertainties) using the voxel-wise minimum for CTV-coverage and voxel-wise maximum for organ-at risk dose distributions Both VMAT and IMPT plans were optimized according to the dose parameters that were included as predictors in the NTCP models of the NIPP such that sparing priority was given to the 5 following DVH-parameters: D_{mean} oral cavity, D_{mean} contralateral parotid, D_{mean} pharyngeal constrictor muscles superior, D_{mean} pharyngeal constrictor muscle inferior, D_{mean} cricopharyngeal muscle, in order to obtain lowest NTCP values possible for that patient [7]. Further details of the NTCP models, treatment and planning were described in previous reports [7,13,14].

We use the same CT and immobilization devices for photon and proton planning. For patients referred from other institutes for proton therapy, a new planning CT is made at our department for IMPT plan creation. We always anticipate for the rare event of machine failure and in that case patients can be treated with the same immobilization on their VMAT backup plan.

Since the Dutch Medical Research Involving Human Subjects Act is not applicable to data collection as part of routine clinical practice, the requirement of written informed consent was waived by the ethics committee. However, all patients are informed upfront that their data can be used for research purposes and are offered the possibility to refuse participation. The data used in this study is a part of 'Standard Follow-up Program (SFP) for Head and Neck Cancer Patients Treated With Curative Primary or Postoperative Radiotherapy or Chemoradiation (H&NTOX)', ClinicalTrials.gov Identifier: NCT02435576.

Basic preselection tool

A VMAT plan was created for all HNC patients as a first step in model-based selection workflow. After completing the VMAT plan, a calculation algorithm, referred to as basic preselection tool, was used in order to select patients for the plan comparison. In this basic preselection tool, the OAR D_{mean} of the clinical VMAT plan and the baseline data of the patient relevant for the NTCP models were used for $\text{NTCP}_{\text{VMAT}}$ calculation, whereas it assumes zero Gy_{RBE} D_{mean} for all OARs for the $\text{NTCP}_{\text{IMPT}}$ calculation. If the ΔNTCP -value between these two NTCP values exceeded at least one of the ΔNTCP -thresholds, a plan comparison was indicated and an IMPT plan was created [7].

Proposed new preselection tool

Instead of assuming a 0 Gy_{RBE} dose for all OARs for the IMPT plan, we created more advanced tools with models to estimate the IMPT D_{mean} in the five OARs in the NTCP models for more accurately identifying HNC patients who may benefit from proton therapy. The workflow of these advanced preselection tools has been shown in the [Fig. 1](#).

In order to predict IMPT D_{mean} of the five OARs, 5 methods were developed as illustrated in [Fig. 2](#). For each OARs, separate predictions were performed using these methods. For each linear regression model used, bootstrapping was performed for internal validation to avoid overfitting, to correct coefficients and model performances for optimism as described in the TRIPOD statement [15]. When a predicted value for an OAR D_{mean} resulted as negative value (< 0 Gy_{RBE}), it was set to 0 Gy_{RBE}.

To this purpose, we used the data of 141 patients who were selected by the basic preselection tool for IMPT planning. In order to obtain VMAT D_{mean} values and volume of the OAR parts for the method 1,2 and 3, we created the following substructures for the OARs ([Fig. 3](#)):

- OAR_inPTV = the overlapping part of the OARs with the PTV_5425 expanded by 5 mm margin, where the dose can be reduced only marginally with IMPT;
- OAR_outPTV = the OAR part outside the PTV_5425 expanded by 5 mm margin, where the dose reduction with IMPT is expected to be more pronounced

The PTV_5425 always encompassed the PTV_7000. The 5-mm margin was given to account for a minimum unavoidable dose fall-off [14].

A case example to illustrate more clearly how to use the five methods are demonstrated in the [supplementary data Appendix B: Case 1](#).

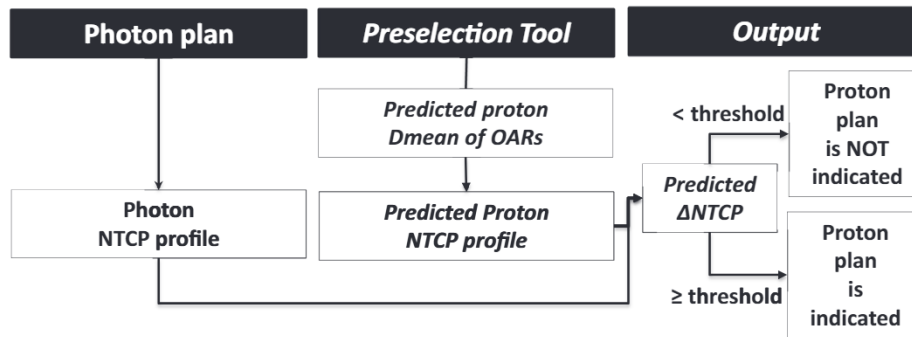


Fig. 1. The preselection tool workflow: The preselection tool 1) predicts the OARs D_{mean} values for an IMPT plan, 2) calculates the NTCP-values using these predicted OARs D_{mean} values, 3) compares the predicted NTCP-values with $NTCP_{VMAT}$ -values, 4) outputs an IMPT plan creation indication only if the predicted $\Delta NTCP$ -values meet the threshold values in the NIPP.

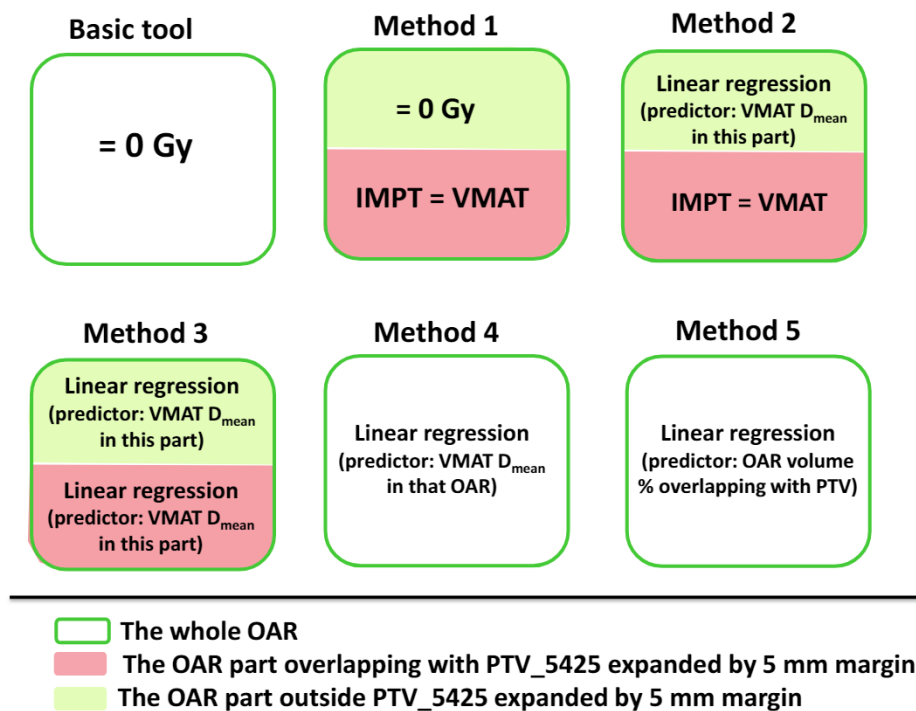


Fig. 2. The assumptions and prediction methods used for IMPT OAR D_{mean} predictions in the basic preselection tool and five methods proposed for the advanced preselection tool. For Methods 1, 2 and 3, separate assumptions and predictions were made for the OAR parts overlapping and outside the PTV_{5425} expanded by 5 mm, the volume-weighted average of OAR parts were taken to calculate D_{mean} for the whole OAR. In Method 4 and 5, the IMPT D_{mean} in the whole OAR was predicted using linear regression model which used $VMAT D_{mean}$ in that OAR and OAR volume % that was overlapping with PTV_{5425} expanded by 5 mm, respectively.

Comparison of the methods

The grouped results for each method were analysed in terms of

- Specificity (% of prevented redundant IMPT plan comparisons),
- Sensitivity (% of accurately detected patients who qualified for proton therapy based on actual plan comparison),
- Accuracy (the overall probability that a patient was correctly classified by the methods in terms of selection for proton therapy).

In addition, the methods were compared in terms of R^2 values and residual values. R^2 values indicate the quality of regression models and the percentage of explained variance with 1 indicating a perfect fit. The limitation of R^2 values is that it does not take into account whether the residuals (the difference between actual and predicted IMPT D_{mean}) are positive or negative. Therefore, methods

were also compared in terms of residual values to better interpret their outcomes using Friedman test. The residual values affect the efficacy and accuracy of a method, since positive (actual $D_{mean} >$ predicted D_{mean}) and negative (actual $D_{mean} <$ predicted D_{mean}) residuals lead to overestimation (decreased specificity) and underestimation (decreased sensitivity) of the IMPT potential in terms of OAR sparing, respectively. In addition, higher absolute residual values indicate larger differences between actual and predicted values, i.e., poor prediction performance.

Post-hoc adjustment of the predicted IMPT D_{mean} values

In case of a false negative selection outcome by a method, the predicted IMPT D_{mean} -values for all OARs using that method were rescaled and reduced by 1% gradually until the false negative rate was set to zero. Rescaling (within the method) was performed to

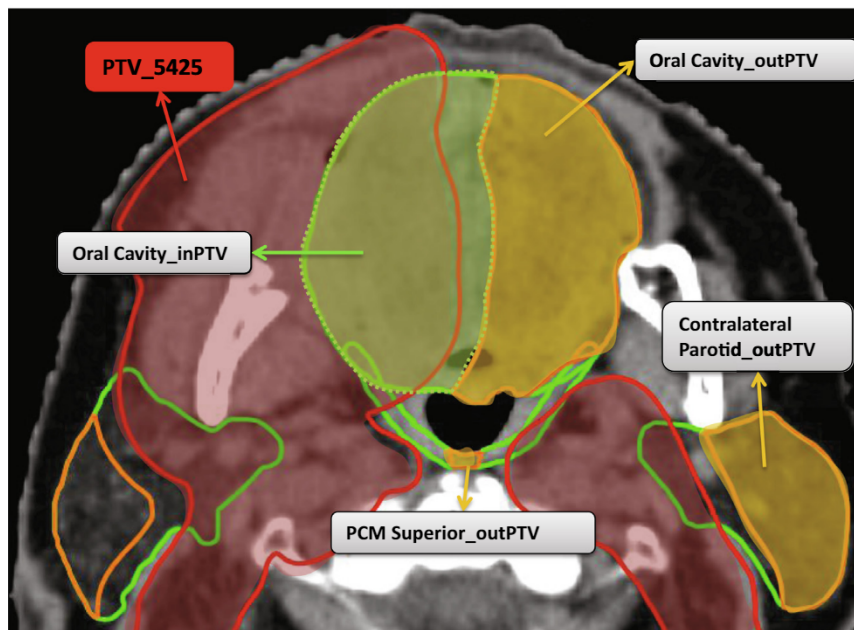


Fig. 3. The substructures of the OARs; OAR_inPTV and OAR_outPTV.

those methods of which the sensitivity rate was below 100%. In order to reduce the predicted values by 1%, they were multiplied by 0.99; to reduce by 2%, they were multiplied by 0.98, etc.

Results

For all 172 patients, a model-based optimized VMAT plan was created and the NTCP-profile was evaluated. In 31 patients (18%), the basic preselection tool was negative such that creating an IMPT plan for comparison was deemed unnecessary. In the remaining 141 patients (82%), a plan comparison was performed. In these patients, a model-based optimized IMPT plan was generated and used to calculate the Δ NTCP-values between photon and proton plan. Of those, 70 (41%) were eventually selected for proton therapy, so the sensitivity and specificity of the basic preselection tool were 100% and 30% while the positive and negative predictive value were 50% and 100%, respectively.

There were 71 patients (41%) with an IMPT plan who were not selected for proton therapy for whom IMPT planning could potentially have been prevented by a more advanced preselection tool.

A strong correlation between the actual and predicted IMPT D_{mean} values (without rescaling) was observed and the 5 methods were comparable with each other with an average R^2 values of 0.927 ranging from 0.904 to 0.942. The predicted NTCP-values highly correlated with the actual $NTCP_{IMPT}$ -values and similar

results across the methods, with an average R^2 value of 0.953 (range: 0.939–0.962) (Table 1, Appendix A: Table A2).

When the methods were compared in terms of the residual values, overall absolute residuals were smaller in Method 2, 3 and 5 compared with Method 1 and 4 ($p < 0.001$) (Fig. 4).

When the methods proposed for the advanced preselection tool were evaluated in terms of diagnostic accuracy measures, the specificity was found to be 49%, 99%, 97%, 96% and 91% for Method 1 to 5, respectively. The sensitivity of the methods was 100%, 16%, 31%, 26% and 63%, respectively; meaning that 0%, 84%, 69%, 74% and 37% of the patients who actually selected for proton therapy would have been wrongfully denied by Method 1 to 5, respectively (Table 2). Therefore, all methods except method 1 required a reduction in predicted values to avoid false negatives. To reach a 100% sensitivity rate, a reduction in the predicted IMPT OAR D_{mean} values by 15%, 16%, 14% and 15% was required for the methods 2 to 5, respectively. With these post-hoc adjustments, the prevented redundant IMPT planning rates (specificity) were 49%, 66%, 68%, 66% and 68% with 70–81% accuracy (% of plan comparisons of which result were correctly classified by the methods) for the method 1 to 5, respectively (Fig. 5, Table 2). A wrongfully denied case example and application of post-hoc adjustment in the predicted values are shown in Appendix B: Case 2.

Fig. 5 shows the specificity, sensitivity and accuracy of the basic preselection tool and the methods per 1% reduction in the pre-

Table 1
The R^2 values of the methods for the OAR D_{mean} and NTCP value predictions for IMPT.

	Method 1	Method 2	Method 3	Method 4	Method 5
R^2 values for IMPT OARs D_{mean} predictions					
Oral Cavity	0.976	0.960	0.957	0.880	0.978
Contralateral Parotid	0.885	0.905	0.903	0.856	0.891
Pharyngeal constrictor muscle Inferior	0.947	0.960	0.960	0.946	0.920
Pharyngeal constrictor muscle Superior	0.967	0.964	0.961	0.941	0.939
Cricopharyngeus	0.887	0.921	0.916	0.897	0.871
R^2 values for $NTCP_{IMPT}$ predictions					
Xerostomia	0.919	0.932	0.930	0.896	0.925
Dysphagia	0.982	0.983	0.982	0.966	0.978
Tube feeding dependence	0.954	0.972	0.967	0.955	0.959

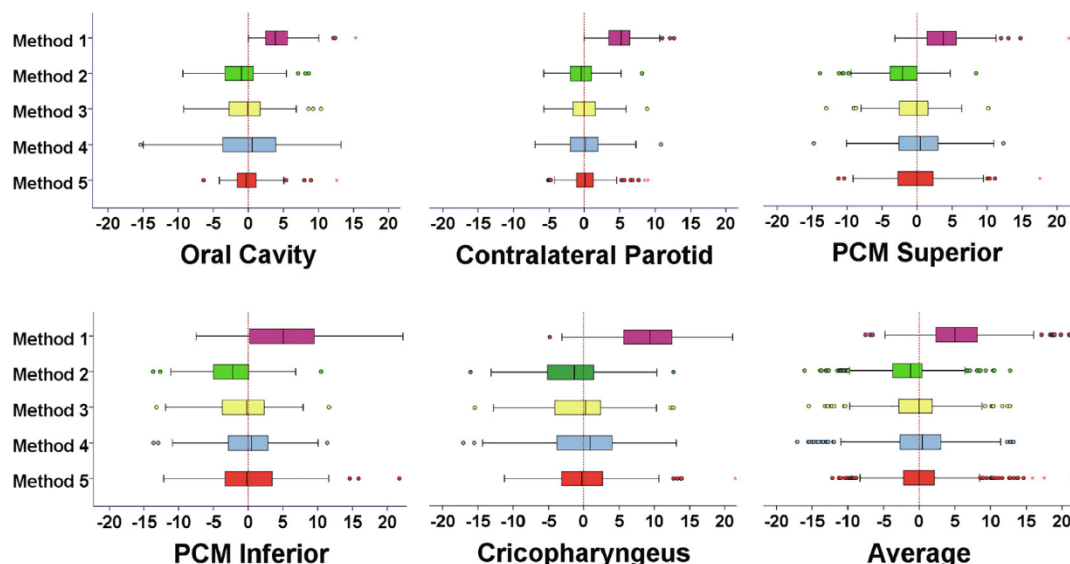


Fig. 4. The boxplots of the residual values for the 5 OARs and their averages for methods 1–5. The positive residual values indicates that actual IMPT dose is higher than the IMPT dose predicted by a given method; i.e., the IMPT potential in terms of OAR sparing was overestimated and may be an indication of IMPT planning by that method for patients who will not qualify for proton therapy (false positivity, less specificity). In contrast, a negative residual value indicates that actual IMPT dose is lower than the predicted value and the IMPT potential is underestimated which may lead to wrongfully denied IMPT planning by that method for patients who qualify for proton therapy (false negative, less sensitivity). In addition, methods with a smaller width box and whiskers (lines extending from the boxes) had a smaller difference between the predicted and actual IMPT dose and the better method performance. For example, from the residual boxplots for oral cavity, it can be seen that the predicted values by Method 1 were smaller than actual IMPT doses (positive residuals) whereas the predicted values by method 2 were generally higher than actual IMPT doses (negative residuals). The difference between predicted and actual doses for the oral cavity were highest and smallest in Method 4 and 5, respectively.

Table 2

The diagnostic performances of the basic PST and the advanced methods with and without reduction factor.

	Basic PST	Method 1	Method 2	Method 3	Method 4	Method 5
<i>The original methods without reduction</i>						
Sensitivity	100% (94.9–100.0)	100.0% (94.9–100.0)	15.7% (8.1–26.4)	31.4% (20.9–43.6)	25.7% (16.0–37.6)	62.9% (50.5–74.1)
Specificity	30.4% (21.7–40.3)	49.0% (39.0–59.1)	99.0% (94.7–100.0)	97.1% (91.6–99.4)	96.1% (90.3–98.9)	91.2% (83.9–95.9)
Positive predictive value	49.7% (46.5–52.9)	57.4% (52.7–62.0)	91.7% (59.23–98.8)	88.0% (69.5–95.9)	81.8% (61.4–92.7)	83.0% (71.9–90.4)
Negative predictive value	100.0% (-)	100.0% (-)	63.1% (60.7–65.5)	67.4% (63.7–70.8)	65.3% (62.0–68.5)	78.2% (72.4–83.0)
Positive likelihood ratio	1.44 (1.3–1.6)	2.0 (1.6–2.4)	16.0 (2.1–121.4)	10.7 (3.3–34.3)	6.6 (2.3–18.6)	7.1 (3.7–13.6)
Negative likelihood ratio	0.0 (-)	0.0 (-)	0.9 (0.8–0.9)	0.7 (0.6–0.8)	0.8 (0.7–0.9)	0.4 (0.3–0.6)
Accuracy	58.7% (51.0–66.2)	69.8% (62.3–76.5)	65.1% (57.5–72.2)	70.4% (62.9–77.1)	67.4% (59.9–74.4)	79.7% (72.9–85.4)
<i>The adjusted methods with reduction in predicted IMPT D_{mean} values</i>						
Sensitivity			100.0% (94.9–100.0)	100.0% (94.9–100.0)	100.0% (94.9–100.0)	100.0% (94.9–100.0)
Specificity			65.7% (55.6–74.8)	67.7% (57.7–76.6)	65.7% (55.6–74.8)	67.7% (57.7–76.6)
Positive predictive value			66.7% (60.5–72.4)	68.0% (61.6–73.7)	66.7% (60.5–72.4)	68.0% (61.6–73.7)
Negative predictive value			100.0% (-)	100.0% (-)	100.0% (-)	100.0% (-)
Positive likelihood ratio			2.9 (2.2–3.8)	3.09 (2.3–4.1)	2.9 (2.2–3.8)	3.09 (2.3–4.1)
Negative likelihood ratio			0.0 (-)	0.0 (-)	0.0 (-)	0.0 (-)
Accuracy			79.7% (72.9–85.4)	80.8% (74.1–86.4)	79.7% (72.9–85.4)	80.8% (74.1–86.4)
Reduction needed			20.0%	16.0%	14.0%	15.0%

95% CI were given in the brackets. The specificity represents the % of prevented redundant IMPT planning whereas sensitivity represents the % of patients detected by methods that actually qualified for proton therapy. As the sensitivity decreases, the % of wrongfully denied patients increases (Sensitivity = 1 – False negative). Positive likelihood ratio = True positive rate/False positive rate. Negative likelihood ratio = False negative rate/True negative rate.

dicted values. Since the sensitivity of the basic preselection tool and method 1 was 100%, no rescaling was required.

Discussion

In our cohort of 141 HNC patients who qualified for a plan comparison, 70 (50%) were selected for proton therapy according to selection criteria defined in NIPP. IMPT plans were also created for the remaining 71 (50%) patients who did not qualify for proton therapy, which were used only for decision making of treatment modality but not for the actual treatment.

Although plan comparisons are essential for model-based selection, redundant IMPT planning has two drawbacks: 1) treatment

delay, which was approximately 2.5 days in our clinic related to IMPT planning but may be even more when patients are referred from other centres; 2) increased costs and extra workload related to IMPT plan preparation, including consultation, simulation and robust IMPT planning and robustness evaluation. These drawbacks emphasize the need for a more efficient selection of patients for IMPT planning using a more sophisticated preselection tool that is able to identify patients with higher probabilities of being selected for proton therapy more accurately.

In the current study, five different methods using a combination of VMAT dose data and data obtained from delineated OARs and target volume in terms of overlap between the two were proposed for a more advanced preselection tool. Each of these have some

Rescaling factor	Specificity						Sensitivity						Accuracy					
	Basic PST	M 1	M2	M3	M4	M5	Basic PST	M 1	M2	M3	M4	M5	Basic PST	M 1	M2	M3	M4	M5
0%	30%	49%	99%	97%	96%	91%	100%	100%	16%	31%	26%	63%	58%	70%	65%	70%	68%	80%
-1%	30%	49%	99%	95%	93%	91%	100%	100%	17%	46%	34%	67%	58%	70%	66%	75%	69%	81%
-2%	30%	49%	99%	93%	91%	90%	100%	100%	20%	54%	36%	71%	58%	70%	67%	77%	69%	82%
-3%	30%	49%	98%	90%	85%	87%	100%	100%	30%	67%	50%	73%	58%	70%	70%	81%	71%	81%
-4%	30%	49%	94%	87%	81%	87%	100%	100%	37%	79%	59%	77%	58%	70%	71%	84%	72%	83%
-5%	30%	49%	92%	83%	77%	87%	100%	100%	50%	86%	69%	83%	58%	70%	75%	84%	74%	85%
-6%	30%	49%	89%	79%	77%	82%	100%	100%	64%	87%	73%	84%	58%	70%	79%	82%	75%	83%
-7%	30%	49%	88%	79%	75%	81%	100%	100%	77%	89%	87%	89%	58%	70%	84%	83%	80%	84%
-8%	30%	49%	84%	78%	73%	80%	100%	100%	83%	91%	89%	91%	58%	70%	84%	83%	80%	84%
-9%	30%	49%	81%	74%	70%	78%	100%	100%	87%	96%	91%	94%	58%	70%	83%	83%	79%	85%
-10%	30%	49%	79%	72%	70%	75%	100%	100%	89%	96%	93%	96%	58%	70%	83%	82%	79%	84%
-11%	30%	49%	78%	71%	69%	75%	100%	100%	91%	97%	97%	97%	58%	70%	83%	82%	80%	84%
-12%	30%	49%	77%	71%	67%	72%	100%	100%	96%	97%	99%	97%	58%	70%	85%	82%	80%	82%
-13%	30%	49%	72%	71%	66%	71%	100%	100%	96%	99%	99%	99%	58%	70%	82%	82%	79%	82%
-14%	30%	49%	71%	70%	66%	70%	100%	100%	97%	99%	100%	99%	58%	70%	82%	82%	80%	82%
-15%	30%	49%	71%	68%	66%	68%	100%	100%	99%	99%	100%	100%	58%	70%	82%	81%	80%	81%
-16%	30%	49%	71%	68%	66%	68%	100%	100%	99%	100%	100%	100%	58%	70%	82%	81%	80%	81%
-17%	30%	49%	69%	68%	66%	68%	100%	100%	99%	100%	100%	100%	58%	70%	81%	81%	80%	81%
-18%	30%	49%	68%	68%	66%	68%	100%	100%	99%	100%	100%	100%	58%	70%	81%	81%	80%	81%
-19%	30%	49%	68%	68%	66%	68%	100%	100%	99%	100%	100%	100%	58%	70%	81%	81%	80%	81%
-20%	30%	49%	66%	68%	66%	68%	100%	100%	100%	100%	100%	100%	58%	70%	80%	81%	80%	81%

Fig. 5. The specificity and sensitivity of the basic preselection tool and the methods proposed for the advanced preselection tool per 1% reduction in the predicted values. The specificity represents the % of prevented redundant IMPT planning whereas sensitivity represents the % of patients detected by methods that actually qualified for proton therapy. As the sensitivity decreases, the % of wrongfully denied patients increases (Sensitivity = 1 - False negative). The accuracy represents the % of the patients that were correctly classified by the methods. The performance of a method increases as the colour changes from red to blue. Of note, no rescaling was further applied when a sensitivity of 100% was achieved for a method.

strengths and limitations. The Method 1 prevented 49% of redundant IMPT planning with no wrongfully denied PT planning for any patients (sensitivity = 100%) and therefore required no post-hoc adjustment. However it suffers from predicting higher IMPT D_{mean} values than actual D_{mean} values, especially in the small OARs such as swallowing muscles, as the dose within the PTV can be lower with IMPT than expected. The method 2 had the lowest sensitivity and highest specificity (without post-hoc adjustment). The method underestimates substantially the potential of IMPT in terms of OAR sparing. The method 3 had higher sensitivity and similar specificity compared with method 2 and 4. The method examines the relationship between IMPT and VMAT D_{mean} values separately in OAR parts overlapping and outside the PTV, provides more information about the dose-fall with protons near and distant to the target areas. The Method 4 had lowest sensitivity following method 2. The method requires no OAR substructures creation in the TPS, thereof can be regarded as easier to use in the clinic. The method 5 had higher sensitivity compared with method 2-4 with an increased specificity from 30% to 91% compared with basic PST without any post-hoc adjustment. The method uses only the volume data of the OARs with regard to their overlap with PTV_5425 and does not require a VMAT plan data as an input, unlike other methods.

The first method (best-case scenario for IMPT plan) was selected as the new advanced preselection tool and replaced the basic preselection tool in our clinic. This method resulted in less redundant plan comparisons with the no risk of false negative outcomes in our cohort. The other 4 methods, using linear regression models to predict the IMPT dose to OARs, bear this risk of false negative results. To avoid false negativity, a reduction in the predicted values was used for. The optimal reduction magnitude can be determined according to institutional preference. Based on institu-

tional predilection and facilities, the issue of wrongfully denied patients can be prioritized and the risk can be minimized using a greater reduction in institutions with sufficient resources, while reducing redundant planning can be a priority in institutions with limited resources where a smaller reduction in the predicted values can be chosen (see Fig. 5).

It is important to note that in all patients included in the current study, model-based optimized plans were created for both VMAT and IMPT [16]. As the quality of the VMAT and IMPT plans, optimization strategies, dose scheduling and patient characteristics are expected to differ widely across centres, the models presented in this study may not be valid for use in other centres, as both regression coefficients of the parameters in the models as well as the level of rescaling is expected to differ from center to center [17-21]. Also within institutions, or specific subgroups of patients, inter-patient variance could be larger and the performance and applicability of any model could be reduced. Therefore, it is essential to assess the model parameters and rescaling factors, by validating, and if necessary revising or updating our models with own institute-specific patient data. Moreover, as radiation technologies and center performance evolve over time, regular updating of the model and rescaling factors is paramount within each centre.

In model-based selection of patients for proton therapy, machine learning methods and knowledge-based dose predictions can also be used to predict plan comparison outcome of new patients by learning from previous ones who have already a plan comparison. A number of authors reported on knowledge-based dose predictions and automated planning for both photons and protons with a high accuracy and strong correlations between predicted and manual dose distributions [17,22-26]. Furthermore, multifactorial decision support systems using continuously

learning artificial intelligence is a promising field to be used for such a purpose [27]. Although these are fast and highly precise methods, they may accommodate the same drawbacks associated with a small dataset lacking in variation, as some of our methods do. Because they remain largely dependent on the plans based on which the models were created and thus are also prone to be affected by the consistency of the plans which is hard to achieve with inter-center and even inter-planner variability [17,28]. Moreover, all models require continuous or regular updating to account for learning curve effects and new technological developments. In order to overcome these obstacles, a large library which is continuously fed and constantly updated with optimal photon and proton plans is required in the context of rapid-learning health care system. In order to provide level 1 evidence to demonstrate the real potential of proton therapy with randomized control trials consisting of true beneficiaries, international collaboration to create such a database is vital [1,2,29–31].

The methods proposed in the current study do not require any infrastructure and are easy to implement and straightforward to use in the clinic. Furthermore, a validation is not required for Method 1 since it is primarily based on the best-case scenario for the IMPT plan. This is also the main reason why it was selected as the new advanced preselection tool in our clinic, as the NIPP has recently been updated with new NTCP models. Our plan optimization strategy in terms of OAR sparing prioritization has been adapted to these new NTCP models, i.e., OARs in these new NTCP models are given priority for sparing, which results in different dose trade-offs between OARs. Consequently, the models created using the plans optimized based on previous NTCP models must be validated for use with plans with optimization based on new NTCP models. The Method 1 is independent from optimization strategies, as it assumes 0 Gy_{RBE} dose for the OAR parts outside the PTV.

Conclusion

Model-based selection of patients for proton therapy can be optimized by selecting patients for plan comparison more effectively with advanced preselection tools that are able to predict the potential outcome of a plan comparison. The five methods proposed in the current study prevent redundant IMPT planning in more than 65% of the patients who would not qualify for PT in our institution, without wrongfully denying any patient proton therapy. Validation of the tools is warranted before clinical application in other centres.

Conflict of interest

None.

Funding

This work described in this paper has been granted by the European Union (Project number: 730983: INSPIRE-project).

Acknowledgement

This work described in this paper has been granted by the European Union (Project number: 730983: INSPIRE-project).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.radonc.2021.04.012>.

References

- Langendijk JA, Lambin P, De Ruyscher D, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. *Radiother Oncol* 2013;107:267–73.
- Langendijk JA, Boersma LJ, Rasch CRN, van Vulpen M, Reitsma JB, van der Schaaf A, et al. Clinical trial strategies to compare protons with photons. *Semin Radiat Oncol* 2018;28:79–87.
- Landelijk Platform Protontherapie (LPPT). Landelijk Indicatieprotocol Protontherapie Hoofdhals. 2017.
- Beetz I, Schilstra C, van Luijk P, Christianen MEMC, Doornaert P, Bijl HP, et al. External validation of three dimensional conformal radiotherapy based NTCP models for patient-rated xerostomia and sticky saliva among patients treated with intensity modulated radiotherapy. *Radiother Oncol* 2012;105:94–100.
- 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.
- Wopken K, Bijl HP, van der Schaaf A, van der Laan HP, Chouvalova O, Steenbakkers RJHM, et al. Development of a multivariable normal tissue complication probability (NTCP) model for tube feeding dependence after curative radiotherapy/chemo-radiotherapy in head and neck cancer. *Radiother Oncol* 2014;113:95–101.
- Tambas M, Steenbakkers RJHM, van der Laan HP, Wolters AM, Kierkels RGJ, Scandurra D, et al. First experience with model-based selection of head and neck cancer patients for proton therapy. *Radiother Oncol* 2020;151:206–13.
- Beddok A, Vela A, Calugaru V, Tessonier T, Kubes J, Dutheil P, et al. Proton therapy for head and neck squamous cell carcinomas: a review of the physical and clinical challenges. *Radiother Oncol* 2020;147:30–9.
- Murphy CT, Galloway TJ, Handorf EA, Egleston BL, Wang LS, Mehra R, et al. Survival impact of increasing time to treatment initiation for patients with head and neck cancer in the United States. *J Clin Oncol* 2016;34:169–78.
- Tribius S, Donner J, Pazdyka H, Münscher A, Gröbe A, Petersen C, et al. Survival and overall treatment time after postoperative radio(chemo)therapy in patients with head and neck cancer. *Head Neck* 2016;38:1058–65.
- Shaikh T, Handorf EA, Murphy CT, Mehra R, Ridge JA, Galloway TJ. The impact of radiation treatment time on survival in patients with head and neck cancer. *Int J Radiat Oncol Biol Phys* 2016;96:967–75.
- Fredriksson A, Forsgren A, Hårdemark B. Minimax optimization for handling range and setup uncertainties in proton therapy. *Med Phys* 2011;38:1672–84.
- 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.
- Kierkels RGJ, Wopken K, Visser R, Korevaar EW, van der Schaaf A, Bijl HP, et al. Multivariable normal tissue complication probability model-based treatment plan optimization for grade 2–4 dysphagia and tube feeding dependence in head and neck radiotherapy. *Radiother Oncol* 2016;121:374–80.
- Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1. <https://doi.org/10.7326/M14-0698>.
- 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.
- Delaney A, 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.
- Verbakel WFA, Doornaert PAH, Raaijmakers CPJ, Bos LJ, Essers M, van de Kamer JB, et al. Targeted intervention to improve the quality of head and neck radiation therapy treatment planning in the Netherlands: short and long-term impact. *Int J Radiat Oncol Biol Phys* 2019;105:514–24.
- Tol JP, Dahele M, Doornaert P, Slotman BJ, Verbakel WFA. Different treatment planning protocols can lead to large differences in organ at risk sparing. *Radiother Oncol* 2014;113:267–71.
- Esposito M, Maggi G, Marino C, Bottalico L, Cagni E, Carbonini C, et al. Multicentre treatment planning inter-comparison in a national context: the liver stereotactic ablative radiotherapy case. *Phys Med* 2016;32:277–83.
- Das IJ, Cheng CW, Chopra KL, Mitra RK, Srivastava SP, Glatstein E. Intensity-modulated radiation therapy dose prescription, recording, and delivery: patterns of variability among institutions and treatment planning systems. *J Natl Cancer Inst* 2008;100:300–7.
- Delaney AR, Dahele M, Tol JP, Kuijper IT, Slotman BJ, Verbakel WFA. Using a knowledge-based planning solution to select patients for proton therapy. *Radiother Oncol* 2017;124:263–70.
- Kierkels RGJ, Fredriksson A, Both S, Langendijk JA, Scandurra D, Korevaar EW. Automated robust proton planning using dose-volume histogram-based mimicking of the photon reference dose and reducing organ at risk dose optimization. *Int J Radiat Oncol Biol Phys* 2019;103:251–8.
- Tol JP, Dahele M, Gregoire V, Overgaard J, Slotman BJ, Verbakel WFA. Analysis of EORTC-1219-DAHANCA-29 trial plans demonstrates the potential of knowledge-based planning to provide patient-specific treatment plan quality assurance. *Radiother Oncol* 2019;130:75–81.
- Krayenbuehl J, Zamburlini M, Ghandour S, Pachoud M, Tanadini-Lang S, Tol J, et al. Planning comparison of five automated treatment planning solutions for locally advanced head and neck cancer. *Radiat Oncol* 2018;13:170.

- [26] 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.
- [27] Walsh S, de Jong EEC, van Timmeren JE, Ibrahim A, Compter I, Peerlings J, et al. Decision support systems in oncology. *JCO Clin Cancer Inform* 2019:1–9.
- [28] 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.
- [29] Langendijk JA, Orecchia R, Haustermans K, Zips D, Balosso J, Lacombe D, et al. Prospective data registration and clinical trials for particle therapy in Europe. *Radiother Oncol* 2018;128:9–13.
- [30] Grau C, Baumann M, Weber DC. Optimizing clinical research and generating prospective high-quality data in particle therapy in Europe: Introducing the European Particle Therapy Network (EPTN). *Radiother Oncol* 2018;128:1–3.
- [31] Grau C, Durante M, Georg D, Langendijk JA, Weber DC. Particle therapy in Europe. *Mol Oncol* 2020;14:1492–9.