



Article

A Decision Support Tool to Optimize Selection of Head and Neck Cancer Patients for Proton Therapy

Makbule Tambas *, Hans Paul van der Laan , Arjen van der Schaaf , Roel J. H. M. Steenbakkers and Johannes Albertus Langendijk

Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, 9713 GZ Groningen, The Netherlands; h.p.van.der.laan@umcg.nl (H.P.v.d.L.); a.van.der.schaaf@umcg.nl (A.v.d.S.); r.steenbakkers@umcg.nl (R.J.H.M.S.); j.a.langendijk@umcg.nl (J.A.L.)

* Correspondence: m.tambas@umcg.nl; Tel.: +31-050-361-6161

Simple Summary: A decision support tool was developed to select head and neck cancer patients for proton therapy. The tool uses delineation data to predict expected toxicity risk reduction with proton therapy and can be used before a treatment plan is created. The positive predictive value of the tool is >90%. This tool significantly reduces delays in commencing treatment and avoid redundant photon vs. proton treatment plan comparison.

Abstract: Selection of head and neck cancer (HNC) patients for proton therapy (PT) using plan comparison (VMAT vs. IMPT) for each patient is labor-intensive. Our aim was to develop a decision support tool to identify patients with high probability to qualify for PT, at a very early stage (immediately after delineation) to avoid delay in treatment initiation. A total of 151 HNC patients were included, of which 106 (70%) patients qualified for PT. Linear regression models for individual OARs were created to predict the D_{mean} to the OARs for VMAT and IMPT plans. The predictors were OAR volume percentages overlapping with target volumes. Then, actual and predicted plan comparison decisions were compared. Actual and predicted OAR D_{mean} (VMAT $R^2 = 0.953$, IMPT $R^2 = 0.975$) and NTCP values (VMAT $R^2 = 0.986$, IMPT $R^2 = 0.992$) were highly correlated. The sensitivity, specificity, PPV and NPV of the decision support tool were 64%, 87%, 92% and 51%, respectively. The expected toxicity reduction with IMPT can be predicted using only the delineation data. The probability of qualifying for PT is >90% when the tool indicates a positive outcome for PT. This tool will contribute significantly to a more effective selection of HNC patients for PT at a much earlier stage, reducing treatment delay.



Citation: Tambas, M.; van der Laan, H.P.; van der Schaaf, A.; Steenbakkers, R.J.H.M.; Langendijk, J.A. A Decision Support Tool to Optimize Selection of Head and Neck Cancer Patients for Proton Therapy. *Cancers* **2022**, *14*, 681. <https://doi.org/10.3390/cancers14030681>

Academic Editors: Lorenzo Preda and David Wong

Received: 27 December 2021

Accepted: 27 January 2022

Published: 28 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: proton therapy; head and neck cancer; plan comparison; IMPT; patient selection; model based selection; decision support tool; dose prediction

1. Introduction

There is a remarkable increase in the number of head and neck cancer (HNC) patients treated with proton therapy (PT) worldwide [1–3]. Based on the 2019 data of the National Association for Proton Therapy survey, HNC patients constituted 14.2% of all patients treated in 28 PT centers in USA, compared to 5.8% in 2012 and showed the strongest increase in the number of patients treated with PT compared with other tumor sites [4]. A similar trend was also observed in European PT centers, with HNC being the most commonly treated indication following CNS tumors and comprising 15% of all adult patients treated with PT last year [5].

One of the main issues in PT application in HNC is how to select patient for PT who are likely to benefit most from PT in terms of toxicity risk reduction compared with photons. Given the fact that early treatment initiation is required for a better survival outcome in HNC [6–8], estimation of the expected benefit from PT as early as possible is clinically

relevant. One way to determine the benefit from PT is to make a photon (volumetric-modulated arc therapy (VMAT)) vs. proton (intensity modulated proton therapy (IMPT)) treatment plan comparison and translate the differences in dose to organs-at-risk (OARs) (i.e., Δ Dose), using normal tissue complication probability (NTCP) models, to expected toxicity risk differences (Δ NTCP) [9,10]. When using multiple NTCP-models, a Δ NTCP profile can be created that can be considered as a biomarker for the expected benefit of protons compared to photons. This called model-based selection has been used in the Netherlands since 2018 to select HNC patients for PT [9,11–13]. However, creating an *in silico* plan comparison for each patient is time consuming and may not be a feasible option for some centers. Centers that do not have a PT facility may refrain from consulting a PT center to check the suitability of a patient to be treated with PT, as this procedure may delay the initiation of treatment. They need to create a VMAT plan first, send it to a PT center combined with other patient data and then have to wait for an IMPT plan to be created and the results of a plan comparison being send back. It might result in treatment delays, especially for patients for whom a plan comparison result shows limited benefit from PT, in addition to the fee which may be asked by the PT center for the comparison that may not be covered by insurance companies.

Therefore, there is an unmet need for decision support tools which can estimate the expected benefit from PT without the need of a full plan comparison. The main advantages of such a tool would be to decrease delays in treatment initiation and to make efficient use of available resources. It may also allow for physicians to discuss the option for PT with their patients based on reliable estimates of the potential gain, which will thereby enhance shared decision making on a referral for PT, and whether this might be worthwhile [14–17].

The aim of this study was to develop a decision support tool to predict the toxicity risk reduction with PT (i.e., the Δ NTCP profile). The tool is to be used immediately after OAR and target volume delineation, but before performing any treatment planning. We aimed to investigate the use of delineation data only to predict both IMPT and VMAT OAR dose profiles. The decision support tool had to be straightforward, easy to implement in routine clinical practice while able to identify patients who are highly likely to gain from PT compared with regular photon treatment.

2. Materials and Methods

Our study comprised 151 patients treated with primary radiotherapy \pm systemic treatment who were subjected to the model-based selection procedure between September 2019 and December 2020.

The OARs and target delineation was performed on the simulation CT according to the international consensus guidelines for CT-based delineation of OARs and targets in the head and neck region, using MRI and PET/CT imaging of the patients [18–21]. The radiotherapy schedule consisted of 54.25 Gy for PTV_5425 and 70 Gy for PTV_7000 in 35 fractions using a simultaneous integrated boost technique for both photons and protons (constant RBE 1.1). The characteristics of the VMAT and IMPT plans were discussed in detail in our previous study [22].

2.1. NTCP Models and Δ NTCP Thresholds

To test if patients qualified for PT, they were evaluated according to the updated Dutch National Indication Protocol for Proton Therapy, which includes four NTCP models, for Grade ≥ 2 xerostomia and dysphagia and Grade ≥ 3 xerostomia and dysphagia. In addition to baseline xerostomia and dysphagia complaints (Appendix A, Table A1), these NTCP models include the D_{mean} of eight OARs as predictors, including the oral cavity, bilateral submandibular and parotid glands and the pharyngeal constrictor muscle (PCM) superior, medius and inferior [23].

The Δ NTCP thresholds for PT selection were as follows: (1) $\geq 10\%$ for Grade ≥ 2 toxicities, (2) $\geq 5\%$ for Grade ≥ 3 toxicities, (3) $\geq 15\%$ for the summed Δ NTCP ($\Sigma\Delta$ NTCP) of

Grade ≥ 2 toxicities (with a minimum of $\geq 5\%$ for each) or (4) $\geq 5\%$ for $\Sigma\Delta\text{NTCP}$ of Grade ≥ 3 toxicities (with a minimum of $\geq 3.75\%$ for each) [23].

2.2. IMPT and VMAT OAR D_{mean} Prediction

Overlap structures were created for each OAR for their part overlapping with PTV_7000 (OAR_in70) and for the OAR part outside the PTV_7000 but inside the PTV_5425 (OAR_in54_out70). Next, we expanded both of the PTVs with different margins including 3, 5, 7, 10 and 15 mm (with an estimated penumbra range of 0–15 mm) and created additional OAR overlap structures with these expanded PTVs (Figure 1).

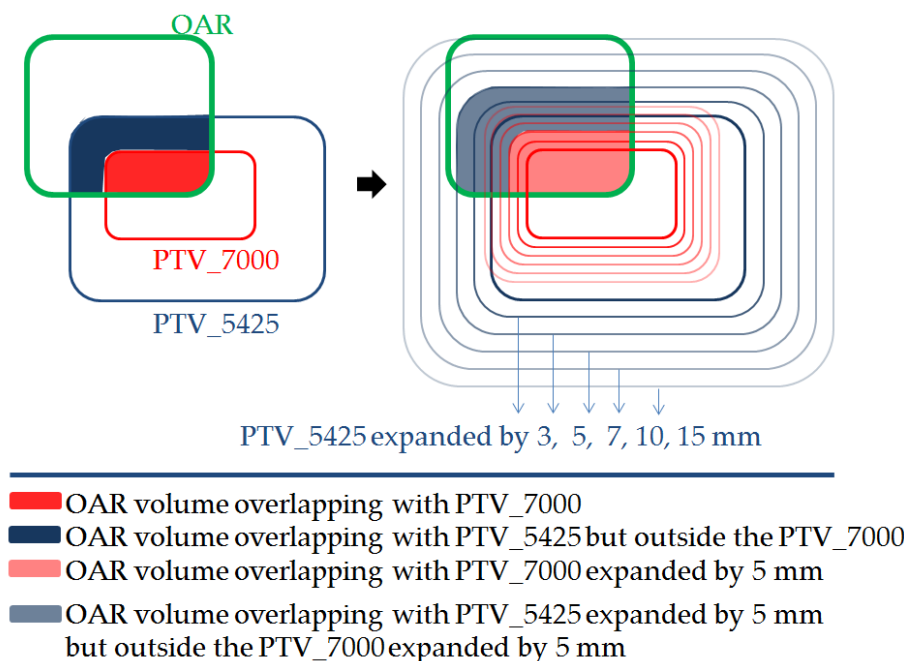


Figure 1. OAR volume overlapping with PTVs and PTVs expanded by different margins.

In order to determine the optimal PTV expansions and corresponding OAR overlap structures for D_{mean} predictions, we created linear regression models for each OAR, where the endpoint was the D_{mean} of that OAR with either VMAT or IMPT. The predictors were the percentages of the OAR_in70 and OAR_in54_out70 volumes of that OARs. In total, 56 linear regression models were created (8 OARs \times 6 different margins (0, 3, 5, 7, 10 and 15 mm) \times 2 modalities (VMAT and IMPT)) to determine the most promising PTV expansion margin for OAR D_{mean} prediction. The performance of the models was evaluated by the goodness-of-fit R^2 values (1 = perfect fit: predicted and actual values are equal). The margin that provided the model with the highest R^2 value was selected for that OAR for VMAT and IMPT predictions.

2.3. The Proposed Decision Support Tool

Using the delineated OARs and target volumes and the overlap between them, the D_{mean} of the eight OARs in the NTCP models were predicted for VMAT and IMPT plans separately. Subsequently, NTCPs and ΔNTCP values were calculated. Based on the ΔNCTP thresholds mentioned above, the predicted plan comparison decision was determined for each patient (Figure 2).

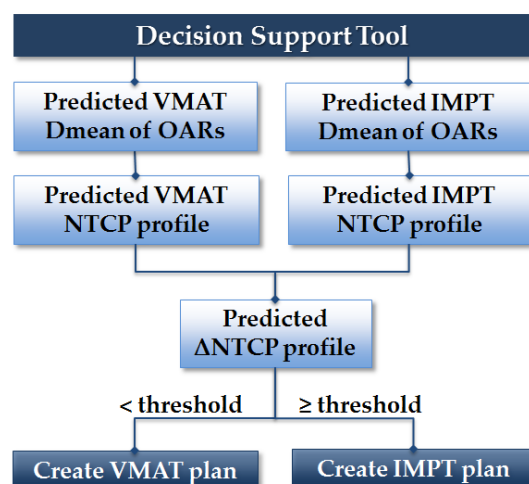


Figure 2. The workflow of the decision support tool.

Diagnostic Measures of the Decision Support Tool

The predicted and actual plan comparison outcomes were compared, and the sensitivity, specificity, positive and negative predictive value and accuracy of the proposed decision support tool were determined. In addition, the robustness of the tool within different patient subgroups (based on treatment initiation date and based on tumor location) were determined and compared with the 95% CI values of the decision support tool in the entire patient cohort for the following two reasons:

- The quality of the plans can be improved when more experience is gained, as there is always a learning curve when a new treatment modality is implemented in a clinic (in this case IMPT) [24–26]. To account for that learning curve, patients were sorted based on their treatment initiation date and the population was divided into two subgroups. First, the initial 70 patients treated and second, the remaining 71 patients, who were treated more recently. Then, the diagnostic measures of the tool within these two subgroups were determined.
- The D_{mean} of the OARs and the frequency of being selected for PT differ based on the primary tumor location, which may also impact the performance of the tool among patients with different tumor locations. In order to examine this, patients were divided into three different groups based on the primary tumor location, i.e., ‘pharynx’, ‘larynx’ and ‘others’. Subsequently, the diagnostic measures of the tool were determined within these three subgroups.

2.4. Statistical Analysis

A chi-squared test was used to examine the differences between groups by baseline categorical characteristics. For continuous variable comparisons, statistical tests were selected based on type (related and independent samples) and distribution (normal and non-normal) of the data. The differences in predicted OAR-doses and NTCP-values for VMAT and IMPT plans were compared using the Wilcoxon Signed Rank Test or the paired samples T-test, whichever appropriate. All statistical tests were two-sided and a p -value of ≤ 0.05 was considered statistically significant. Analyses were performed using the Statistical Package for Social Sciences (SPSS) for Windows, version 21.0 (SPSS Inc., Chicago, IL, USA).

3. Results

3.1. Patient and Selection for Proton Therapy

Of the 151 patients included in this study, 106 (70%) patients qualified for PT, while 45 (30%) patients did not and were treated with VMAT. Patients’ characteristics are shown in the Appendix A, Table A2.

For the different tumor locations, PT selection rates were different and the dominant NTCP model that triggered patient selection differed per location. In general, most patients were selected based on dysphagia related models or based on the $\Sigma\Delta$ NTCP of grade ≥ 2 toxicities. However, laryngeal cancer patients were mainly selected based on xerostomia grade ≥ 2 models (Appendix A, Table A3).

3.2. VMAT and IMPT OAR D_{mean} Prediction Results

The R^2 values of the 56 linear regression models created to predict VMAT and IMPT OAR D_{mean} are shown in Figure 3. The R^2 values of the models for IMPT OAR D_{mean} predictions were generally higher than those for VMAT. The R^2 values of the best performing linear regression models varied from 0.903 to 0.954 for VMAT and from 0.946 to 0.985 for IMPT.

PTV expansion margin	0 mm	3 mm	5 mm	7 mm	10 mm	15 mm
R^2 values for the VMAT D_{mean} predictions						
Oral Cavity	0.834	0.861	0.874	0.883	0.895	0.903
PCM Superior	0.866	0.901	0.921	0.937	0.950	0.913
PCM Medius	0.785	0.869	0.908	0.938	0.939	0.856
PCM Inferior	0.849	0.904	0.927	0.946	0.954	0.902
Parotid Left	0.763	0.883	0.919	0.930	0.924	0.888
Parotid Right	0.810	0.890	0.907	0.912	0.903	0.844
Submandibular Left	0.704	0.821	0.875	0.911	0.918	0.838
Submandibular Right	0.709	0.819	0.870	0.899	0.907	0.868
R^2 values for the IMPT D_{mean} predictions						
Oral Cavity	0.965	0.978	0.982	0.985	0.985	0.971
PCM Superior	0.935	0.964	0.974	0.972	0.952	0.873
PCM Medius	0.881	0.938	0.949	0.944	0.896	0.757
PCM Inferior	0.912	0.949	0.963	0.966	0.949	0.864
Parotid Left	0.830	0.933	0.960	0.965	0.952	0.901
Parotid Right	0.888	0.950	0.962	0.960	0.938	0.857
Submandibular Left	0.845	0.926	0.946	0.944	0.900	0.734
Submandibular Right	0.873	0.946	0.963	0.956	0.911	0.815

Figure 3. Assessment of the best performing PTV expansion margin to predict IMPT and VMAT OAR D_{mean} . The R^2 values are given with the highest R^2 values for each OAR are outlined in red. The values become higher as the color of the cells gets darker.

Selected PTV Expansion Margins for D_{mean} Predictions

In general, 10 mm (for VMAT) and 5 mm (for IMPT) PTV expansion margins resulted in models with the highest R^2 values (Figure 3). Eventually, for VMAT D_{mean} predictions, a 10 mm PTV expansion margin was selected for five of the eight OARs (PCM superior, PCM medius, PCM inferior and left and right submandibular gland), 7 mm PTV expansion margins were selected for two OARs (left and right parotid) and 15 mm for one OAR (oral cavity). On the other hand, for IMPT D_{mean} predictions, a 5 mm PTV expansion margin was selected for five of the eight OARs (PCM superior, PCM medius, right parotid and left and right submandibular glands) and 7 mm for three OARs (oral cavity, PCM inferior and left parotid gland). The coefficients of the selected models are given in the Appendix A, Table A4.

3.3. VMAT and IMPT NTCP Prediction Results

The predicted and actual NTCP values highly correlated with R^2 values of 0.986 for VMAT and 0.992 for IMPT (Figure 4).

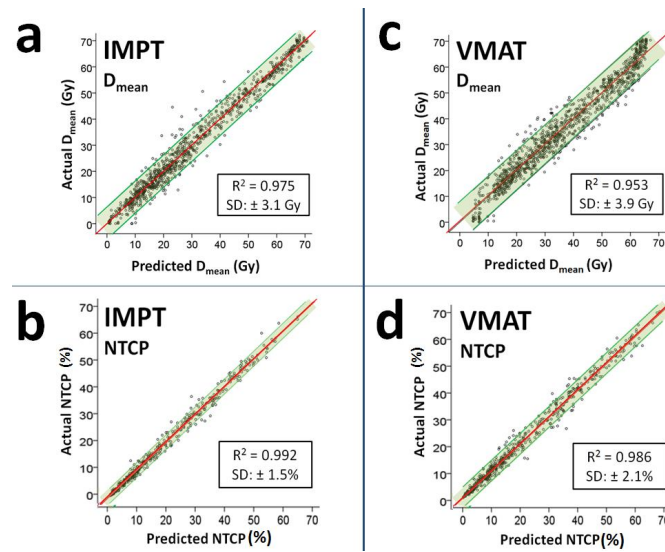


Figure 4. Predicted vs. actual values for eight OAR D_{mean} for IMPT (a) and VMAT (c); and predicted vs. actual four NTCP values for IMPT (b) and VMAT (d) shown in the same scatter plot. The red lines are the origins where predicted and actual values are equal to each other. The green areas indicate the 95% CIs for the individual predictions.

The residual values, i.e., the average \pm SD differences between predicted and actual NTCP values for VMAT and IMPT, were $0.4\% \pm 2.6\%$ and $0.1\% \pm 1.7\%$ for Grade ≥ 2 dysphagia, $0.4\% \pm 2.0\%$ and $0.2 \pm 0.8\%$ for Grade ≥ 3 dysphagia, $0.3 \pm 2.2\%$ and $0.0 \pm 2.2\%$ for Grade ≥ 2 xerostomia and $0.1 \pm 0.9\%$ and $0.0 \pm 0.9\%$ for Grade ≥ 3 xerostomia, respectively (Figure 5).

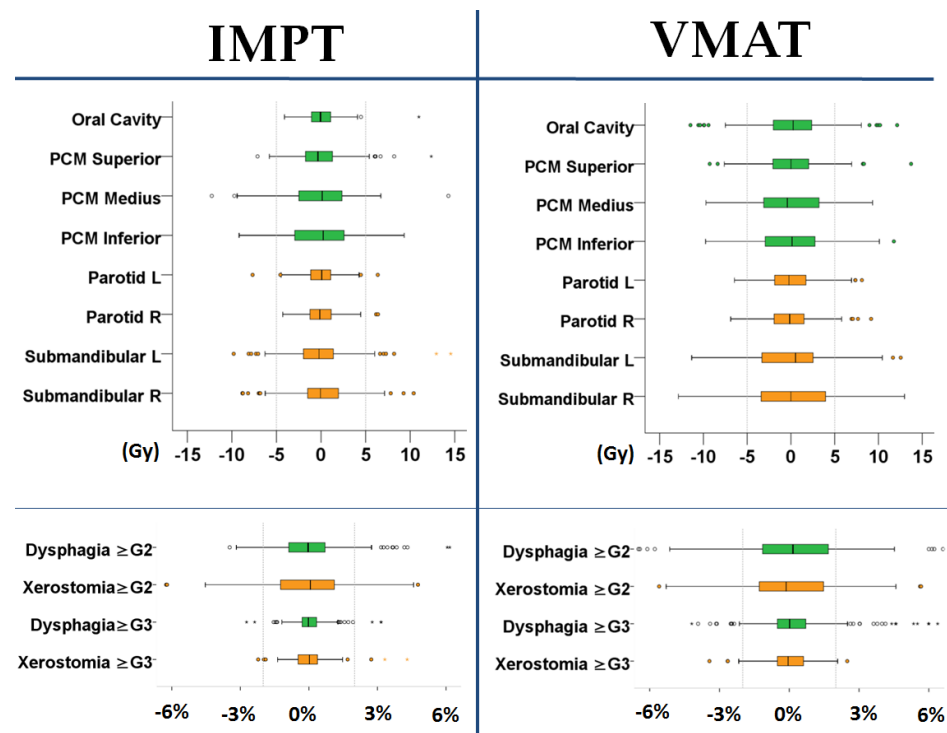


Figure 5. The boxplots of the residuals, i.e., the difference between predicted and actual OAR D_{mean} (upper) and NTCP values (lower) for IMPT and VMAT. Dysphagia- and xerostomia-related OARs and NTCPs are shown in green and orange, respectively. Dots in the figure represent outliers, i.e., values more than 1.5 interquartile range (IQR) but less than 3 IQR from the end of the box).

3.4. Diagnostic Measures of the Decision Support Tool

The sensitivity and specificity of the decision support tool were 64% (95% CI: 54–73) and 87% (95% CI: 73–95), respectively. The positive and negative predictive value and accuracy were 92% (95% CI: 84–96), 51% (95% CI: 44–58) and 71% (95% CI: 63–78), respectively. The post-hoc sensitivity analysis revealed that the diagnostic measures of the tool were within 95% CI limits among the first and second half of the patient population (Figure 6).

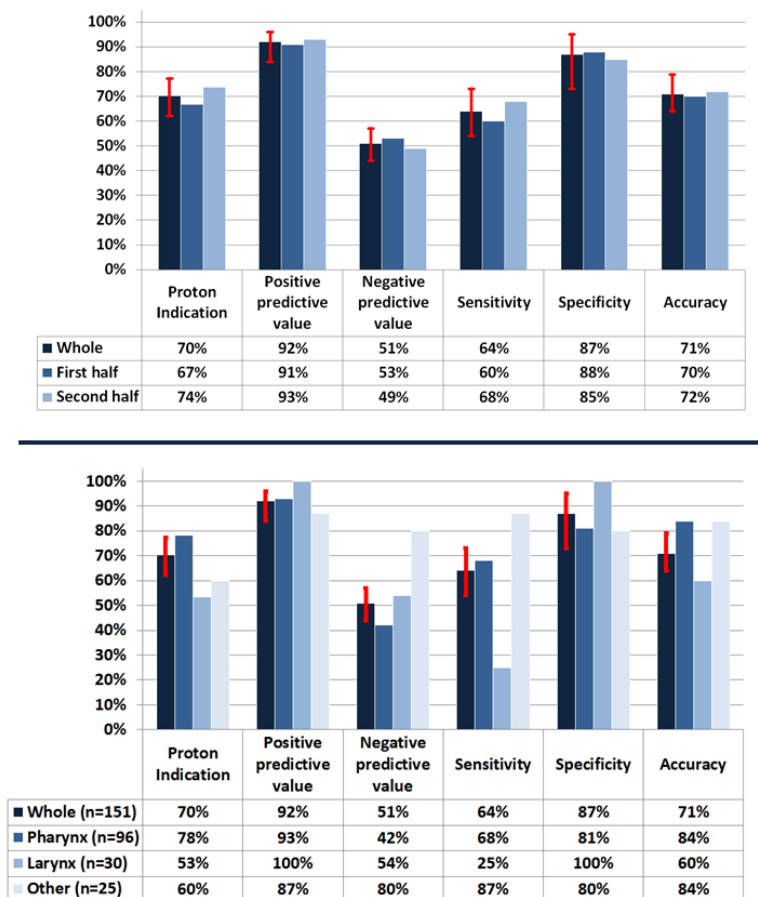


Figure 6. Variability of diagnostic measures across split cohort groups (upper) and across tumor locations (lower). The red bars on the columns of the whole patient cohort indicate 95% CI of the decision support tool.

When the diagnostic measures of the decision support tool were evaluated based on the tumor location, a high variability was observed in terms of negative predictive value and sensitivity (Figure 6). In particular, sensitivity was low for the laryngeal tumors compared with other tumor locations.

When we compared the actual VMAT plan (instead of the predicted VMAT plan) with the predicted IMPT plan, the diagnostic measures of the decision support tool increased from 64 to 77% for sensitivity, 84 to 89% for specificity, 92 to 94% for positive predictive value, 51 to 63% for negative predictive value and 71 to 81% for accuracy.

4. Discussion

In this study, we developed a decision support tool to select patients for either VMAT or IMPT treatment based on the delineated OARs and PTVs before treatment planning. In our patient group, the positive predictive value of the tool was >90%, indicating that patients had a very high probability to be selected for PT in case of a positive decision support tool outcome. The negative predictive value of the tool was 51%, meaning that there is still approximately 50% probability that the patient will qualify for PT when the

tool results a negative indication. The main advantage of such a tool is to identify HNC patients, at a very early stage of the preparation phase of radiotherapy, that are most likely to qualify for PT. The high positive predictive value of our tool can provide this information with a high level of confidence to both physicians and patients. In the Netherlands, in case of a positive outcome of the tool for PT, a plan comparison still has to be made as this is required to get reimbursement.

In our previous study, different versions of a pre-selection tool were proposed that required VMAT planning first or used a single overlap definition, only for PTV_5425 and which used the same margin for all OARs [27]. We now improved our models compared to our previous study. We predicted the Δ NTCP profile solely based on delineated structures without any treatment planning and determined the optimal PTV expansion margin for each OAR that could be used in conjunction with the percentage OAR overlap to predict both VMAT and IMPT OAR D_{mean} , without the need to first perform VMAT planning.

To increase the level of confidence in the probability of being selected for PT, the NTCP profile of a VMAT plan created by the referring center could be compared with the predicted IMPT NTCP profile by the decision support tool for an even higher level of confidence, owing to the fact that the VMAT prediction is the main source of uncertainty and the correlation between the actual and predicted NTCP_{IMPT} values is relatively high ($R^2 = 0.992$ vs. 0.946). Indeed, when we compared the actual VMAT plan (instead of the predicted VMAT plan) with the predicted IMPT plan, the performance of the decision support tool further improved.

Since the sensitivity and negative predictive value of the decision support tool is relatively low, institution-specific changes can be made to increase its sensitivity, i.e., to decrease false negative cases, depending on available resources, treatment capacity and preferences of that institution. In PT centers with sufficient resources where a high sensitivity is the priority, three different adjustments can be applied: (1) to rescale the predicted IMPT D_{mean} values using a uniform or OAR-specific rescaling factor, (2) to use either the lower (for IMPT) or upper (for VMAT) bound 95 CI% of the D_{mean} prediction models coefficients and (3) to decrease predicted IMPT D_{mean} values using the SD values of the residuals. This might be especially needed for patients with laryngeal cancer. The sensitivity of the decision support tool is lowest for this group, which can be explained by the relatively lower correlation between actual and predicted D_{mean} values for salivary glands compared with other OARs, given that laryngeal cancer patients qualify for PT mostly based on xerostomia-related Δ NTCP thresholds (Appendix A, Table A3). When these three proposed post-hoc adjustments were applied to the predicted D_{mean} values for salivary glands, the performance of the decision support tool for laryngeal cancer patients changed as follows: (1) when predicted IMPT D_{mean} was rescaled by 0.85, sensitivity increased from 25% to 63%, while specificity decreased from 100% to 50%, (2) when the lower bounds of the 95% CI of the coefficients were used to predict IMPT D_{mean} , sensitivity was 57%, and specificity was 57% and (3) when predicted IMPT D_{mean} values were decreased by 1 SD of the residuals (3.1 Gy, see Figure 4), sensitivity was 81%, and specificity was 50%. If a sensitivity of 100% is the priority for a given institute, then a greater post-hoc adjustment can be applied, at the cost of plan comparison for more patients that would not eventually be selected for PT.

Automated planning combined with machine learning approaches can also be used to preselect patients for plan comparison based on delineation [28–37]. A recent study by Kouwenberg et al. investigated the potential of using automated planning in combination with machine learning to be used for preselection in 45 HNC patients who were subjected to model-based selection based on a previous version of the Dutch national indication protocol [37]. They compared the actual photon plan with an IMPT plan that was generated with non-clinical, fully-automated planning. Similar to our study, when the Δ NTCP thresholds were directly applied, it led to false negative and positive outcomes with an overall accuracy of 82%. To reach a sensitivity of 100% with minimal false positive results, the machine learning (Gaussian naïve Bayes classifier) was used to define the optimal

decision boundary. It increased the sensitivity to 100% with an overall accuracy of 87%. Nevertheless, both their and our studies emphasize two main conclusions: (1) D_{mean} of the OARs can be predicted with a high accuracy either using automated planning or using a simpler approach of the OAR and PTV geometric relationship in terms of overlap; (2) using the predicted values directly leads to false positive and false negative results due to variation between manual and predicted doses, which can be overcome by using either advanced artificial intelligence methods as proposed by Kouwenberg et al. or using simpler post-hoc adjustments in the predicted values, as proposed in the current study.

The method proposed in the current study has some limitations, which are inherent to single center planning comparison study and, thus, the generalizability of the results in another settings. First, patients used in this study were treated with only two prescribed dose levels for the PTV (i.e., 54.25 and 70 Gy). Second, we used model-based optimization, in which the OARs were prioritized during optimization based on the NCTP gain per Gy increase in their D_{mean} values. Thus, it is very likely that the coefficients of the models used for the decision support tool are different in other centers where different dose schedules or other planning priorities are used [38]. Therefore, refitting the decision support tool models is warranted before clinical application in any center that would like to use it. Furthermore, the models need to be constantly updated as the expertise and optimization strategies evolve in time within a center and may differ from center to center. Moreover, technological developments in either photon or PT (e.g., dynamic arc proton therapy) over time may improve dose conformity and may jeopardize the performance of the tool, requiring further adjustments. Lastly, we currently present the diagnostic measures of the decision support tool based on the NTCP models and ΔNTCP thresholds as they are defined in the Dutch National Indication Protocol for Proton Therapy [23]. Different thresholds or protocols may be used to select HNC patients for PT in other countries.

5. Conclusions

We developed a decision support tool to select patients for PT that can be used before any treatment plan is created. The expected toxicity reduction with PT can be predicted using only the delineation data. The probability of qualifying for PT is >90% when the tool indicates a positive outcome for PT based on 5% and 10% NTCP reduction thresholds for grade ≥ 2 and ≥ 3 dysphagia and xerostomia, respectively. This tool can contribute significantly to identifying HNC patients at a much earlier stage that are highly likely to benefit from PT. The tool avoids clinical workload, is cost effective and can be used without delaying treatment initiation.

Author Contributions: Conceptualization, M.T., H.P.v.d.L. and J.A.L.; methodology, M.T., H.P.v.d.L., A.v.d.S. and J.A.L.; software, M.T. and H.P.v.d.L.; validation, M.T., H.P.v.d.L. and A.v.d.S.; formal analysis, M.T. and H.P.v.d.L.; investigation, M.T. and H.P.v.d.L.; resources, M.T., H.P.v.d.L. and J.A.L.; data curation, M.T.; writing—original draft preparation, M.T.; writing—review and editing, M.T., H.P.v.d.L., A.v.d.S., R.J.H.M.S. and J.A.L.; visualization, M.T.; supervision, H.P.v.d.L., R.J.H.M.S. and J.A.L.; project administration, J.A.L.; funding acquisition, J.A.L. All authors have read and agreed to the published version of the manuscript.

Funding: The research described in this paper has been funded by the European Union (Project number: 730983: INSPIRE-project).

Institutional Review Board Statement: 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 were informed upfront that their data can be used for research purposes and were offered the possibility to refuse participation. The data used in this study are a part of the research project ‘Standard Follow-up Program (SFP) for Head and Neck Cancer Patients Treated with Curative Primary or Postoperative Radiotherapy or Chemoradiation (H&NTOX)’, [ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT02435576) Identifier: NCT02435576.

Informed Consent Statement: 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 were informed upfront that their data can be used for research purposes and were offered the possibility to refuse participation.

Data Availability Statement: This study was based on the data derived from the prospective data registration program from department of Radiation Oncology of the UMCG. The data are part of a much larger project and will be used for a number of future research projects. Consequently, to present the data cannot be shared.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. NTCP models used for patients selection in revised Dutch National Indication Protocol for Proton Therapy.

Variables	Endpoint (6 Months after Radiotherapy)			
	Xerostomia		Dysphagia	
	Grade ≥ 2	Grade ≥ 3	Grade ≥ 2	Grade ≥ 3
Constant (B0)	−2.2951	−3.7286	−4.0536	−7.6174
$\sqrt{D_{\text{mean}}}$ Parotid ipsilateral + $\sqrt{D_{\text{mean}}}$ Parotid contralateral	0.0996	0.0855		
D_{mean} submandibular bilateral	0.0182	0.0156		
D_{mean} Oral cavity			0.0300	0.0259
D_{mean} PCM superior			0.0236	0.0203
D_{mean} PCM medius			0.0095	0.0303
D_{mean} PCM inferior			0.0133	0.0341
Baseline xerostomia: None (EORTC QLQ-H&N35—Q41: score 1)	0.0000	0.0000		
Baseline xerostomia: A littler (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
Tumor location (Pharynx)			−0.6281	0.0387
Tumor location (Larynx)			−0.7711	−0.5303

Table A2. Patient characteristics and comparison of the first and second half of the patients based on their treatment initiation date.

		First Half n (%)	Second Half n (%)	Total n (%)	p Value
Tumor Location	Oropharynx	28 (37)	33 (43)	61 (40)	0.131
	Larynx	21 (28)	9 (12)	30 (20)	
	Hypopharynx	7 (9)	12 (16)	19 (13)	
	Nasopharynx	8 (11)	8 (11)	16 (11)	
	Oral cavity	6 (8)	11 (14)	17 (11)	
	Other	5 (7)	3 (4)	8 (5)	
Baseline Xerostomia	None	35 (47)	44 (58)	79 (52)	0.141
	A little	28 (37)	27 (36)	55 (36)	
	Quite	12 (16)	5 (7)	17 (11)	
Baseline Dysphagia	None	52 (69)	54 (71)	106 (70)	0.795
	Grade 2	22 (29)	20 (26)	42 (28)	
	Grade 3–5	1 (1)	2 (3)	3 (2)	
Proton Indication	No	25 (33)	20 (26)	45 (30)	0.346
	Yes	50 (67)	56 (74)	106 (70)	
	Total	75 (100)	76 (100)	151 (100)	

Table A3. PT selection frequency based on tumour location and NTCP models. Note that some patients qualified for protons based on more than one criteria.

	Nasopharynx	Oral Cavity	Oropharynx	Hypopharynx	Larynx	Other	Total
Dysphagia grade ≥ 2	44%	59%	44%	37%	7%	25%	36%
$\Sigma\Delta$ NTCP of grade ≥ 2	50%	12%	38%	32%	17%	0%	29%
Dysphagia grade ≥ 3	19%	18%	18%	63%	3%	0%	20%
Xerostomia grade ≥ 2	19%	0%	3%	0%	33%	0%	10%
Xerostomia grade ≥ 3	6%	0%	0%	0%	7%	0%	2%
$\Sigma\Delta$ NTCP of grade ≥ 3	6%	0%	0%	0%	0%	0%	1%
Total	94%	76%	75%	74%	53%	25%	70%

Table A4. The coefficients of the models for VMAT and IMPT D_{mean} predictions.

		B	Std. Error	p Value
Coefficients for VMAT OAR D_{mean} Prediction				
Oral Cavity	Constant (B0)	11.923	0.748	<0.001
	% of Oral Cavity volume overlapping with PTV70 + 15mm	0.447	0.016	<0.001
	% of Oral Cavity volume overlapping with PTV54 + 15 mm but outside PTV70 + 15 mm	0.338	0.062	<0.001
PCM_Superior	Constant (B0)	9.826	0.887	<0.001
	% of PCM_Sup volume overlapping with PTV70 + 10 mm	0.558	0.011	<0.001
	% of PCM_Sup volume overlapping with PTV54 + 10 mm but outside PTV70 + 10 mm	0.382	0.017	<0.001
PCM_Medius	Constant (B0)	4.723	1.196	<0.001
	% of PCM_Med volume overlapping with PTV70 + 10 mm	0.596	0.013	<0.001
	% of PCM_Med volume overlapping with PTV54 + 10 mm but outside PTV70 + 10 mm	0.405	0.019	<0.001
PCM_Inferior	Constant (B0)	4.681	1.180	<0.001
	% of PCM_Inf volume overlapping with PTV70 + 10 mm	0.607	0.013	<0.001
	% of PCM_Inf volume overlapping with PTV54 + 10 mm but outside PTV70 + 10 mm	0.380	0.023	<0.001
Parotid Left	Constant (B0)	6.737	0.538	<0.001
	% of Parotid_left volume overlapping with PTV70 + 7 mm	0.558	0.016	<0.001
	% of Parotid_left volume overlapping with PTV54 + 7 mm but outside PTV70 + 7 mm	0.497	0.029	<0.001
Parotid Right	Constant (B0)	7.205	0.597	<0.001
	% of Parotid_right volume overlapping with PTV70 + 7 mm	0.598	0.016	<0.001
	% of Parotid_right volume overlapping with PTV54 + 7 mm but outside PTV70 + 7 mm	0.466	0.031	<0.001
Submandibular Left	Constant (B0)	5.499	1.292	<0.001
	% of Submand_left volume overlapping with PTV70 + 10 mm	0.568	0.014	<0.001
	% of Submand_left volume overlapping with PTV54 + 10 mm but outside PTV70 + 10 mm	0.471	0.019	<0.001
Submandibular Right	Constant (B0)	6.801	1.415	<0.001
	% of Submand_right volume overlapping with PTV70 + 10 mm	0.566	0.016	<0.001
	% of Submand_right volume overlapping with PTV54 + 10 mm but outside PTV70 + 10 mm	0.437	0.021	<0.001
Coefficients for IMPT OAR D_{mean} Prediction				
Oral Cavity	Constant (B0)	1.481	0.290	<0.001
	% of Oral Cavity volume overlapping with PTV70 + 7 mm	0.641	0.016	<0.001
	% of Oral Cavity volume overlapping with PTV54 + 7 mm but outside PTV70 + 7 mm	0.558	0.043	<0.001
PCM_Superior	Constant (B0)	6.442	0.606	<0.001
	% of PCM_Sup volume overlapping with PTV70 + 5 mm	0.643	0.009	<0.001
	% of PCM_Sup volume overlapping with PTV54 + 5 mm but outside PTV70 + 5 mm	0.468	0.015	<0.001
PCM_Medius	Constant (B0)	9.890	0.863	<0.001
	% of PCM_Med volume overlapping with PTV70 + 5 mm	0.597	0.012	<0.001
	% of PCM_Med volume overlapping with PTV54 + 5 mm but outside PTV70 + 5 mm	0.393	0.019	<0.001
PCM_Inferior	Constant (B0)	3.952	0.992	<0.001
	% of PCM_Inf volume overlapping with PTV70 + 7 mm	0.641	0.011	<0.001
	% of PCM_Inf volume overlapping with PTV54 + 7 mm but outside PTV70 + 7 mm	0.373	0.026	<0.001

Table A4. Cont.

		B	Std. Error	p Value
Parotid Left	Constant (B0)	1.121	0.467	0.018
	% of Parotid_left volume overlapping with PTV70 + 7 mm	0.606	0.012	<0.001
	% of Parotid_left volume overlapping with PTV54 + 7 mm but outside PTV70 + 7 mm	0.513	0.024	<0.001
Parotid Right	Constant (B0)	3.063	0.433	<0.001
	% of Parotid_right volume overlapping with PTV70 + 5 mm	0.639	0.013	<0.001
	% of Parotid_right volume overlapping with PTV54 + 5 mm but outside PTV70 + 5 mm	0.602	0.030	<0.001
Submandibular Left	Constant (B0)	10.117	0.923	<0.001
	% of Submand_left volume overlapping with PTV70 + 5 mm	0.571	0.012	<0.001
	% of Submand_left volume overlapping with PTV54 + 5 mm but outside PTV70 + 5 mm	0.450	0.017	<0.001
Submandibular Right	Constant (B0)	8.808	0.794	<0.001
	% of Submand_right volume overlapping with PTV70 + 5 mm	0.585	0.010	<0.001
	% of Submand_right volume overlapping with PTV54 + 5 mm but outside PTV70 + 5 mm	0.465	0.016	<0.001

References

- Kim, J.K.; Leeman, J.E.; Riaz, N.; McBride, S.; Tsai, C.J.; Lee, N.Y. Proton Therapy for Head and Neck Cancer. *Curr. Treat. Options Oncol.* **2018**, *19*, 28. [CrossRef]
- Li, X.; Lee, A.; Cohen, M.A.; Sherman, E.J.; Lee, N.Y. Past, present and future of proton therapy for head and neck cancer. *Oral Oncol.* **2020**, *110*, 104879. [CrossRef]
- Frank, S.J.; Blanchard, P.; Lee, J.J.; Sturgis, E.M.; Kies, M.S.; Machtay, M.; Vikram, B.; Garden, A.S.; Rosenthal, D.I.; Gunn, G.B.; et al. Comparing Intensity-Modulated Proton Therapy with Intensity-Modulated Photon Therapy for Oropharyngeal Cancer: The Journey From Clinical Trial Concept to Activation. *Semin. Radiat. Oncol.* **2018**, *28*, 108–113. [CrossRef]
- NAPT Annual Survey Results—2019 Data. Available online: <https://members.proton-therapy.org/page/AnnualSurvey> (accessed on 27 September 2021).
- Tambas, M.; van der Laan, H.; Steenbakkens, R.; Doyen, J.; Timmermann, B.; Orlandi, E.; Hoyer, M.; Haustermans, K.; Georg, P.; Burnet, N.; et al. PH-0328 Current practice for selection of adult patients for proton therapy across Europe. *Radiother. Oncol.* **2021**, *161*, S240–S241. [CrossRef]
- Ang, K.; Trotti, A.; Brown, B.W.; Garden, A.; Foote, R.L.; Morrison, W.H.; Geara, F.B.; Klotch, D.W.; Goepfert, H.; Peters, L.J. Randomized trial addressing risk features and time factors of surgery plus radiotherapy in advanced head-and-neck cancer. *Int. J. Radiat. Oncol.* **2001**, *51*, 571–578. [CrossRef]
- Murphy, C.T.; Galloway, T.J.; Handorf, E.A.; Egleston, B.; Wang, L.S.; Mehra, R.; Flieder, D.B.; Ridge, J.A. Survival Impact of Increasing Time to Treatment Initiation for Patients with Head and Neck Cancer in the United States. *J. Clin. Oncol.* **2016**, *34*, 169–178. [CrossRef]
- Graboyes, E.M.; Kompelli, A.R.; Neskey, D.M.; Brennan, E.; Nguyen, S.; Sterba, K.R.; Warren, G.W.; Hughes-Halbert, C.; Nussenbaum, B.; Day, T.A. Association of Treatment Delays with Survival for Patients With Head and Neck Cancer: A Systematic Review. *JAMA Otolaryngol. Head Neck Surg.* **2019**, *145*, 166–177. [CrossRef]
- Langendijk, J.A.; 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*, 267–273. [CrossRef]
- van de Water, T.A.; Bijl, H.P.; Schilstra, C.; Pijls-Johannesma, M.; Langendijk, J.A. 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*, 366–377. [CrossRef]
- Langendijk, J.A.; Boersma, L.J.; Rasch, C.R.; van Vulpen, M.; Reitsma, J.B.; van der Schaaf, A.; Schuit, E. Clinical Trial Strategies to Compare Protons With Photons. *Semin. Radiat. Oncol.* **2018**, *28*, 79–87. [CrossRef]
- Landelijk Platform Protonentherapie (LPPT). Landelijk Indicatieprotocol Protonentherapie Hoofdhals. 2017. Available online: <http://www.nvro.nl/publicaties/rapporten.html> (accessed on 22 September 2021).
- Langendijk, J.A.; Hoebens, F.J.P.; de Jong, M.A.; Doornaert, P.; Terhaard, C.H.J.; Steenbakkens, R.J.H.M.; Hamming-Vrieze, O.; van de Kamer, J.B.; Verbakel, W.F.A.R.; Keskin-Cambay, F.; et al. National Protocol for Model-Based Selection for Proton Therapy in Head and Neck Cancer. *Int. J. Part. Ther.* **2021**, *8*, 354–365. [CrossRef] [PubMed]
- Shen, C.J.; Hu, C.; Ladra, M.; Narang, A.K.; Pollack, C.E.; Terezakis, S.A. Socioeconomic factors affect the selection of proton radiation therapy for children. *Cancer* **2017**, *123*, 4048–4056. [CrossRef] [PubMed]
- McClelland, S., 3rd; Mitin, T.; Jaboin, J.J.; Ciporen, J.N. Impact of Travel Distance on Radiation Treatment Modality for Central Nervous System Disease. *J. Neurosci. Rural. Pract.* **2019**, *10*, 606–607. [PubMed]
- Liu, S.M.; Brooks, E.D.; Rubin, M.L.; MD Anderson Cancer Center Radiation Oncology Fellows Consortium; Grosshans, D.R.; Frank, S.J.; McAleer, M.F.; McGovern, S.L.; Paulino, A.C.; Woodhouse, K.D. Referral Patterns and Treatment Delays in Medulloblastoma: A Large Academic Proton Center Experience. *Int. J. Part. Ther.* **2021**, *7*, 110. [CrossRef] [PubMed]
- Tarnasky, A.M.; Olivere, L.A.; Ledbetter, L.; Tracy, E.T. Examining the Effect of Travel Distance to Pediatric Cancer Centers and Rurality on Survival and Treatment Experiences: A Systematic Review. *J. Pediatr. Hematol.* **2021**, *43*, 159–171. [CrossRef] [PubMed]

18. Brouwer, C.L.; Steenbakkers, R.J.; Bourhis, J.; Budach, W.; Grau, C.; Grégoire, V.; van Herk, M.; Lee, A.; Maingon, P.; Nutting, C.; et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. *Radiother. Oncol.* **2015**, *117*, 83–90. [[CrossRef](#)]
19. Grégoire, V.; Evans, M.; Le, Q.-T.; Bourhis, J.; Budach, V.; Chen, A.; Eisbruch, A.; Feng, M.; Giral, J.; Gupta, T.; et al. Delineation of the primary tumour Clinical Target Volumes (CTV-P) in laryngeal, hypopharyngeal, oropharyngeal and oral cavity squamous cell carcinoma: AIRO, CACA, DAHANCA, EORTC, GEORCC, GORTEC, HKNPCSG, HNCIG, IAG-KHT, LPRHHT, NCIC CTG, NCRI, NRG Oncology, PHNS, SBRT, SOMERA, SRO, SSHNO, TROG consensus guidelines. *Radiother. Oncol.* **2018**, *126*, 3–24. [[CrossRef](#)]
20. Lee, A.W.; Ng, W.T.; Pan, J.J.; Poh, S.S.; Ahn, Y.C.; AlHussain, H.; Corry, J.; Grau, C.; Grégoire, V.; Harrington, K.J.; et al. International guideline for the delineation of the clinical target volumes (CTV) for nasopharyngeal carcinoma. *Radiother. Oncol.* **2018**, *126*, 25–36. [[CrossRef](#)]
21. Grégoire, V.; Ang, K.; Budach, W.; Grau, C.; Hamoir, M.; Langendijk, J.A.; Lee, A.; Le, Q.T.; Maingon, P.; Nutting, C.; et al. Delineation of the neck node levels for head and neck tumors: A 2013 update. DAHANCA, EORTC, HKNPCSG, NCIC CTG, NCRI, RTOG, TROG consensus guidelines. *Radiother. Oncol.* **2014**, *110*, 172–181. [[CrossRef](#)]
22. Tambas, M.; Steenbakkers, R.J.; van der Laan, H.P.; Wolters, A.M.; Kierkels, R.G.; Scandurra, D.; Korevaar, E.W.; Oldehinkel, E.; van Zon-Meijer, T.W.; Both, S.; et al. First experience with model-based selection of head and neck cancer patients for proton therapy. *Radiother. Oncol.* **2020**, *151*, 206–213. [[CrossRef](#)]
23. Landelijk Indicatie Protocol Protontherapie (versie 2.2) (LIPPv2.2) HOOFD-HALSTUMOREN. 2019. Available online: https://www.nvro.nl/images/documenten/rapporten/2019-08-15__Landelijk_Indicatieprotocol_Protontherapie_Hoofdhals_v2.2.pdf (accessed on 22 September 2021).
24. Singhal, M.; Vincent, A.; Simoneaux, V.; Johnstone, P.A.; Buchsbaum, J.C. Overcoming the Learning Curve in Supine Pediatric Proton Craniospinal Irradiation. *J. Am. Coll. Radiol.* **2012**, *9*, 285–287. [[CrossRef](#)] [[PubMed](#)]
25. Jacobs, M.; Kerkmeijer, L.; de Ruyscher, D.; Brunenberg, E.; Boersma, L.; Verheij, M. Implementation of MR-linac and proton therapy in two radiotherapy departments in The Netherlands: Recommendations based on lessons learned. *Radiother. Oncol.* **2021**, *167*, 14–24. [[CrossRef](#)] [[PubMed](#)]
26. Bentzen, S.M. Randomized controlled trials in health technology assessment: Overkill or overdue? *Radiother. Oncol.* **2008**, *86*, 142–147. [[CrossRef](#)] [[PubMed](#)]
27. Tambas, M.; van der Laan, H.P.; Rutgers, W.; van den Hoek, J.G.M.; Oldehinkel, E.; Meijer, T.W.H.; van der Schaaf, A.; Scandurra, D.; Free, J.; Both, S.; et al. Development of advanced preselection tools to reduce redundant plan comparisons in model-based selection of head and neck cancer patients for proton therapy. *Radiother. Oncol.* **2021**, *160*, 61–68. [[CrossRef](#)]
28. Hall, D.C.; Trofimov, A.V.; Winey, B.A.; Liebsch, N.J.; Paganetti, H. Predicting Patient-specific Dosimetric Benefits of Proton Therapy for Skull-base Tumors Using a Geometric Knowledge-based Method. *Int. J. Radiat. Oncol.* **2017**, *97*, 1087–1094. [[CrossRef](#)]
29. Purdie, T.; Dinniwel, R.E.; Letourneau, D.; Hill, C.; Sharpe, M.B. Automated Planning of Tangential Breast Intensity-Modulated Radiotherapy Using Heuristic Optimization. *Int. J. Radiat. Oncol.* **2011**, *81*, 575–583. [[CrossRef](#)] [[PubMed](#)]
30. Moore, K.L.; Brame, R.S.; Low, D.A.; Mutic, S. Experience-Based Quality Control of Clinical Intensity-Modulated Radiotherapy Planning. *Int. J. Radiat. Oncol.* **2011**, *81*, 545–551. [[CrossRef](#)] [[PubMed](#)]
31. Appenzoller, L.M.; Michalski, J.M.; Thorstad, W.L.; Mutic, S.; Moore, K.L. Predicting dose-volume histograms for organs-at-risk in IMRT planning. *Med. Phys.* **2012**, *39*, 7446–7461. [[CrossRef](#)]
32. Zhu, X.; Ge, Y.; Li, T.; Thongphiew, D.; Yin, F.F.; Wu, Q.J. A planning quality evaluation tool for prostate adaptive IMRT based on machine learning. *Med. Phys.* **2011**, *38*, 719–726. [[CrossRef](#)]
33. Shiraishi, S.; Moore, K.L. Knowledge-based prediction of three-dimensional dose distributions for external beam radiotherapy. *Med. Phys.* **2016**, *43*, 378. [[CrossRef](#)]
34. Yuan, L.; Ge, Y.; Lee, W.R.; Yin, F.F.; Kirkpatrick, J.; Wu, Q.J. Quantitative analysis of the factors which affect the interpatient organ-at-risk dose sparing variation in IMRT plans. *Med. Phys.* **2012**, *39*, 6868–6878. [[CrossRef](#)] [[PubMed](#)]
35. Delaney, A.R.; Dahele, M.; Tol, J.P.; Kuijper, I.T.; Slotman, B.J.; Verbakel, W.F. Using a knowledge-based planning solution to select patients for proton therapy. *Radiother. Oncol.* **2017**, *124*, 263–270. [[CrossRef](#)] [[PubMed](#)]
36. Delaney, A.R.; Verbakel, W.F.; Lindberg, J.; Koponen, T.K.; Slotman, B.J.; Dahele, M. Evaluation of an Automated Proton Planning Solution. *Cureus* **2018**, *10*, e3696. [[CrossRef](#)] [[PubMed](#)]
37. 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*, 224–229. [[CrossRef](#)]
38. Wang, Y.; Breedveld, S.; Heijmen, B.; Petit, S.F. Evaluation of plan quality assurance models for prostate cancer patients based on fully automatically generated Pareto-optimal treatment plans. *Phys. Med. Biol.* **2016**, *61*, 4268–4282. [[CrossRef](#)]