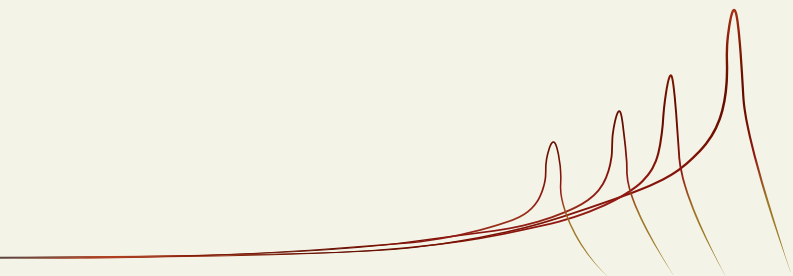
Automated proton therapy planning: clinical validation for head-and-neck cancer and chordoma

Huiskes, M.



Chapter 4

Qualitative Blinded Clinical Assessment of Automated and Manual IMPT Plans for Head and Neck Cancer

Huiskes, M., Astreinidou, E., de Jong, M., den Haan, N., Kroesen, M., Crama, K., Kong, W., Breedveld, S., Heijmen, B., & Rasch, C.

International Journal of Radiation Oncology, Biology, Physics (2025)
https://doi.org/10.1016/J.IJROBP.2025.11.029

ABSTRACT

Purpose: To clinically validate fully automated intensity-modulated proton therapy (IMPT) planning for head and neck cancer (HNC) through blinded comparison with manual plans by experienced radiation oncologists.

Methods: Thirty HNC patients treated with a manually created IMPT plan were retrospectively included. Dose prescription was 7000 cGy(RBE) to the primary tumor and 5425 cGy(RBE) to the bilateral elective nodal volumes. Fully automated IMPT plans were generated using Erasmus-iCycle, applying the same clinical four-beam configuration and dose constraints. Three experienced HNC radiation oncologists (ROs) independently assessed the automated and corresponding manual plans in a blinded manner. Each RO evaluated both plans, blinded to the planning method, to determine clinical acceptability and indicate their preference, if any. Interobserver differences in preference were evaluated using pairwise McNemar tests. A generalized linear mixed model was used to assess overall preference for automated plans.

Results: All 30 automated IMPT plans were considered clinically acceptable, whereas 28 of the 30 manual plans were rated clinically acceptable by all three ROs. In 28 out of 30 cases, the ROs unanimously ($p=1.000$ between each pair of observers) preferred the automated plan. In the remaining 2 cases, one of the three ROs had no preference. The manual plan was never preferred. The overall preference for automated plans over no preference was statistically significant ($p<.001$). Target coverage was considered equivalent, while target conformality and sparing of both serial and parallel OARs were generally rated superior in the automated plans.

Conclusions: All fully automated IMPT plans for HNC were considered clinically acceptable by experienced HNC radiation oncologists. Automated plans were unanimously preferred over manually created plans in nearly all cases, with statistical significance, highlighting the clinical potential of automated IMPT planning tools.

INTRODUCTION

Radiation therapy (RT) treatment planning aims to achieve adequate target coverage while respecting dose constraints for organs at risk (OARs). Manual treatment planning typically requires trade-offs between clinical objectives and involves multiple optimization rounds, including manual adjustment of weights and/or the addition of optimization objectives. Especially in head and neck cancer (HNC), where multiple targets are located close to multiple OARs, manual treatment planning is complex and requires a high level of planner intervention, which makes it a workload and time-consuming process. In addition, the resulting plan quality depends on the experience of the involved planner, and the amount of time available for the planning process.[1,2] These limitations highlight the need for more efficient and consistent methods of treatment plan optimization[3], leading to the development of automated planning systems.

Both in-house developed and commercially available planning solutions have been proposed within the last few years.[4–8] Previously performed validation studies into automated planning have reported successful results for several cancer types, including HNC.[8–11] Most of these studies have been conducted for photon therapy, although several have recently become available for HNC proton therapy.[12–17] The validation in these studies primarily focused on quantitative dosimetric plan metrics, such as dose-volume histogram parameters and clinical objectives. However, for successful integration into clinical practice, both quantitative and qualitative analyses should be conducted.[18]

For automated photon therapy, a few qualitative investigations have been reported in the literature. In a prospective study by Voet et al.[19], treating physicians selected the automatically generated plan over the manual plan in 97% of the HNC cases. Similarly, for prostate cancer, clinicians' preference for automated plans was in line with the observed overall superior dosimetric plan parameters.[20] However, in a multicenter study by Fiandra et al.[21], autoplanning provided significant dosimetric advantages for breast cancer, yet clinicians preferred the automated plans in only half of the cases. Similarly, in lung cancer patients, despite significant OAR dose reductions, the automated plans were not always preferred over their manual counterparts.[22] These observations highlight the importance of qualitative clinical evaluations, even when quantitative dosimetric results indicate superiority of automated plans over manual plans.

However, studies assessing the clinical evaluation of autoplanning for proton therapy remain limited. To date, only one blinded clinical validation study has

investigated deep learning-based planning for intensity-modulated proton therapy (IMPT) in HNC.[23] In contrast to our study, however, the automated plans were not generated fully automatically, as manual adjustments were applied to the deep learning-generated plans.

We previously reported on dosimetric comparisons demonstrating the quality of fully automated IMPT plans for HNC patients.[17] In the present study, we investigated their clinical acceptability and preference through a blinded comparison with manual plans. Radiation oncologists independently assessed and selected between the two, aiming to provide a qualitative evaluation of fully automated IMPT plans and their potential for use in the clinic.

METHODS

Clinical Planning

Thirty-three patients (31 oropharynx, 2 hypopharynx), clinically treated with curative intent in HollandPTC between March 2021 and August 2022, were retrospectively included. All patients were irradiated on a Varian ProBeam unit (Varian Medical Systems, Palo Alto, CA) using a manually created IMPT treatment plans constructed in RayStation v10B (RaySearch Laboratories AB). Informed consent was obtained from all patients for the retrospective use of their clinical data. The prescribed dose was 7000 cGy(relative biologic effectiveness [RBE] = 1.1) to the primary clinical target volume (CTV7000) and 5425 cGy(RBE) to the bilateral elective CTV (CTV5425) in 35 fractions. The plans were constructed by an experienced planner with a 4-beam configuration; gantry angles B1:50°, B2:150°, B3:210°, B4:310°, and a 5cm range shifter. B1 and B4 could be adjusted by the involved planner by 5° to 10° to avoid irradiation through dental fillings. Additionally, avoidance volumes were created to prevent spots passing through dental fillings. Spot traversal through the shoulders was prevented using avoidance structures encompassing the caudal parts of CTV5425. Robust optimization was applied to the CTVs for 21 scenarios with a 3mm isotropic setup uncertainty and ±3% density uncertainty.[24] The plans were optimized in accordance to the planning goals outlined in Table 1, where the dose to OARs was minimized. The clinical IMPT plans were optimized and calculated using the RayStation Monte Carlo dose engine with 1.0% statistical uncertainty.

Table 1 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 inferior part)

* Also in the vw-min dose

** Also in the vw-max dose

Automated planning

The fully automated IMPT plans in this study were generated without human intervention, using multi-criterial optimization as implemented in Erasmus-iCycle. [5,25,26] The optimizer is driven by a treatment site-specific wish-list, which contains hard planning constraints and prioritized objectives. These objectives are sequentially optimized according to their assigned priorities, ensuring that constraints are not violated. After each objective minimization, the achieved result is converted into a constraint, such that its value is maintained during the subsequent optimization of lower-priority objectives. If an objective cannot be fully achieved, a relaxation is applied to enable room for improvement of lower-priority objectives. After objectives were optimized to their goal values, a secondary optimization round is performed to further reduce dose where feasible. The resulting plans are Pareto optimal.[5,25]

The initial definition of the wish-list was based on the clinical target and OAR goals (Table 1) and discussions with clinicians. Of the 33 included patients, three oropharynx cases were randomly selected for iterative wish-list refinement (see Table 1 in the paper of Huiskes et al.[17] for patient and tumor characteristics).

For these cases, automated IMPT plans were evaluated and the wish-list was adjusted until no further improvements were considered achievable. A detailed description of the tuning process can be found in the appendix of Heijmen et al.[20]. The finalized wish-list was subsequently applied to the remaining 30 patients, who were not used in the wish-list development process. These 30 patients and the wish-list were identical to those used in the previous validation study [17], which focused on dosimetric plan comparison.

For all automated plans, the same avoidance structures and monitor unit constraints as in the clinical plans were used, and robust optimization was performed with the same uncertainty scenarios as clinically applied. Dose calculations were conducted with the Astroid dose algorithm [27], configured for the Varian ProBeam beam characteristics [28].

Blind clinician scoring

Three radiation oncologists with clinical experience in treating HNC patients with proton therapy, were individually asked to review the blinded automated and manual clinical treatment plans. All dose data were imported in DICOM RT format into the research version of RayStation 2023B. For each patient, the automated and manual plans were randomly assigned as "Plan A" and "Plan B" to ensure blinding. Consistent with clinical practice, the clinicians assessed both the nominal dose distributions and the robustness evaluations using the voxelwise minimum (vw-min) and voxelwise maximum (vw-max) doses calculated from 28 scenarios with 3mm setup error and ±3% density uncertainty on the planning computed tomography (CT).[29] In addition, the clinical goal template was available to facilitate the evaluation.

Per patient, each RO filled in a digital questionnaire, see Appendix figure 1. In this questionnaire, the clinicians were asked whether each plan was clinically acceptable or not, and whether they preferred Plan A, Plan B, or had no preference. In case of a preference, they could indicate whether they had a low- or high preference towards the preferred plan. Furthermore, they were asked to compare the plans in terms of target coverage, OAR sparing, hotspots, and target conformity, and to identify whether one plan performed better in any of these aspects. Clinicians were also given the opportunity to provide additional free-text comments in the questionnaire.

Statistics

Plan acceptability

Fisher-Exact test was performed to assess whether differences between observers in clinical acceptability of the plans were statistically significant ($p<.05$).

Plan preference

A Generalized Linear Mixed Model was conducted with preference as the dependent variable (0 = no preference, 1 = preference for automated plan), and patient as a random effect to account for the non-independence of observations within patients. The model included only an intercept to test whether there was a statistically significant ($p<.05$) overall preference across all observers.

To investigate interobserver differences between the ROs in plan preference, pairwise comparisons between the observers (RO1, RO2, RO3) were assessed using McNemar test. Plan preferences were initially restructured into 3 categories: 'Manual plan better', 'No preference' and 'Auto plan better'. Since no ratings were given for 'Manual plan better', only the 'No preference' and 'Auto plan better' categories were included in the analysis. This test is appropriate for paired nominal data and evaluates differences in the marginal proportions, taking the paired observation into account.

RESULTS

The manual plans were considered clinically acceptable in 28/30 of the cases, with no statistically significant difference between the observers ($p=0.326$), whereas all 30 automated plans were considered clinically acceptable by the three radiation oncologists. One RO assessed a manual plan as clinically unacceptable due to the presence of a dose bath caudal to the pharyngeal constrictor muscles and larynx, see Figure A2. Another manual plan was considered clinically unacceptable by 1 RO due to an average CTV dose that was judged to be too high.

The 3 ROs had a statistically significant ($p<.001$) overall plan preference towards the automated plan (28/30 patients), with either low- or high impact, while in the remaining 2 patients 1 of the 3 ROs had no preference, see Figure 1 and Table A1. In none of the cases the manual plan was preferred.

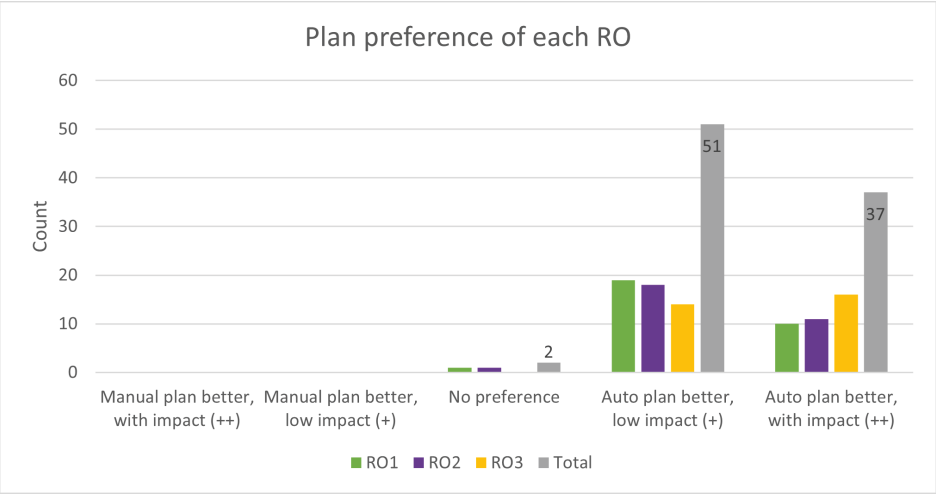


Figure 1 Plan preferences of each radiation oncologist (RO), along with the combined total preferences across all 3 ROs.

The McNemar test revealed no statistically significant differences between any pair of observers (p=1.000 for all comparisons) in plan preference. This result is consistent with the high agreement observed, as nearly all cases were rated by each RO as "Auto plan better" (which includes both 'Auto plan better, low impact +' and 'Auto plan better, with impact ++' categories).

Figure 2 presents the cumulative preference of the 3 observers across subcategories, including target coverage, OAR sparing, hotspots, and target conformity. Overall, target coverage for both CTVs was rated with no preference between the manual and automated plans. However, the automated plans clearly scored better for target conformity and both serial and parallel OAR sparing. Regarding hotspots, most plans were considered equivalent, with a subset demonstrating preference towards the manual- or automated plans. Figure 3 shows an example of dose distributions with blinded naming: 'Plan A' (clinical plan) and 'Plan B' (automated plan). In the case shown, target coverage was rated equivalent, while the automated plan was rated superior by all 3 ROs in terms of OAR sparing and target conformality.

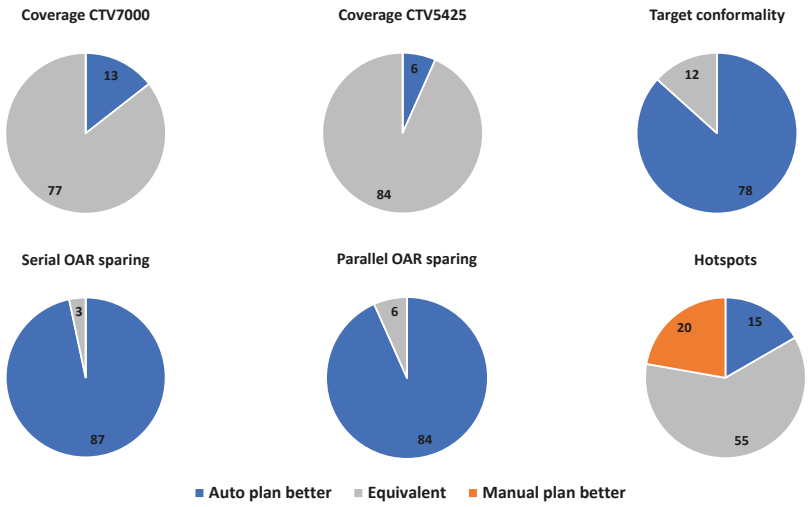


Figure 2 Cumulative preferences of radiation oncologists for automated plans, neither, or manual plans, grouped by subcategory. Each subcategory comprises n = 90 plan comparisons. For target coverage and serial OARs, both nominal and vw-min/vw-max doses were evaluated. Abbreviations: CTV = clinical target volume; OARs = organs at risk; vw-max = voxel wise maximum; vw-min = voxel wise minimum.

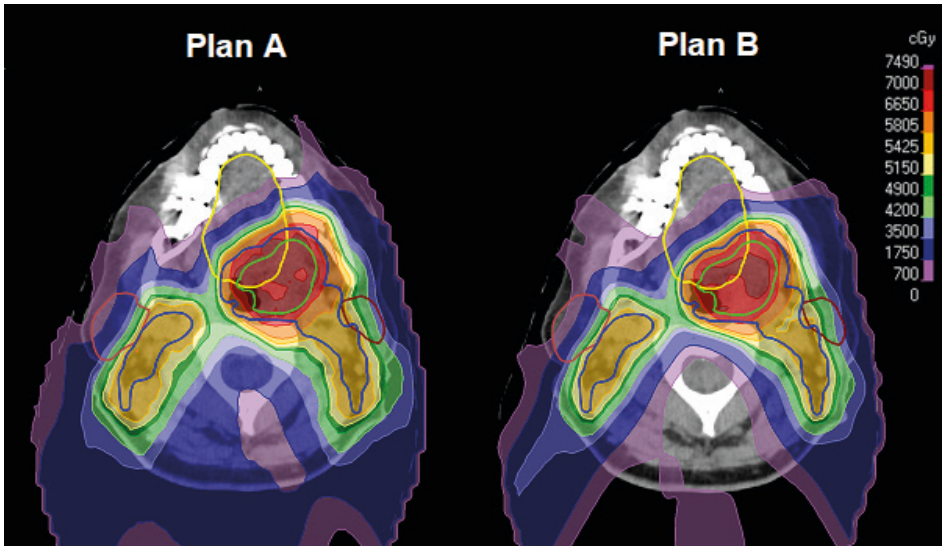


Figure 3 Transversal CT slices showing the nominal dose distribution of "Plan A" (clinical plan) and "Plan B" (automated plan) for a representative case. Structures are labeled as follows: green = CTV7000, dark blue = CTV5425, yellow = oral cavity, and light/dark brown=parotid glands. Target coverage was rated equivalent, whereas Plan B was considered superior in terms of OAR sparing and target conformity. Abbreviations: CT = computed tomography; CTV = clinical target volume; OARs = organs at risk.

In the additional free-text fields, 2 ROs reported about higher mean dose in the oral cavity or cricopharyngeal inlet in 4 of the automated plans, where the other OARs were clearly spared. In 3 automated plans, 1 RO reported about dose through sinuses or shoulders which was less emphasized in the manual plans, and for 1 automated plan, this was addressed by 2 ROs. One RO noted in 3 cases that the average dose to the elective CTV in the automated plan was relatively high, but still within clinical limits.

DISCUSSION

This study presents the first blinded comparison between fully automatically generated and manually constructed IMPT plans for HNC, demonstrating that all automated plans were clinically acceptable and in nearly all cases the automated plans were preferred by the expert ROs. Target coverage was considered equivalent, whereas target conformality and sparing of both serial and parallel OARs were generally superior in the automated plans. These findings demonstrate that fully automated IMPT planning with Erasmus-iCycle produces high-quality, clinically acceptable plans that are consistently favored over manual plans, while also reducing hands-on planning time. This supports the potential for clinical adoption of fully automated IMPT planning as an efficient, high-quality alternative to manual treatment planning.

The results of this study are consistent with our previous work focused on dosimetric parameters, in which we observed comparable target coverage and statistically significant reductions in OAR doses with the automatically generated plans.[17] The superior results of automated planning are obtained through wish-list driven systematic plan generation with automated MCO, which is difficult to reproduce consistently with manual planning. With Erasmus-iCycle, generated plans are guaranteed to be Pareto optimal [5], and all objectives are minimized to the maximum extent while never violating constraints and while respecting objective priorities. Consequently, the optimizer reduces OAR doses wherever feasible without deteriorating higher-priority objectives. In this way, not only are the objective goals met, but OAR doses are also maximally reduced.

Other validation studies of automated planning methods also primarily rely on dosimetric comparisons.[12–16] While dosimetric analysis provides valuable quantitative insights, it does not capture the full spectrum of clinical considerations that treating physicians evaluate during plan review. A major strength of our study is that it advances previous work by incorporating a retrospective, blinded evaluation

of the plans by treating physicians. This approach enables the assessment of qualitative factors that may influence clinical decision-making.

Another strength of this study is the inclusion of a larger number of cases, combined with evaluation by three independent radiation oncologists. In contrast, previous HNC studies have typically assessed fewer cases or involved a smaller number of observers. For example, Voet et al.[19] included 20 cases, while the study by Van Bruggen et al.[23] assessed only 10 to 15 cases; both studies involved a single treating physician, which may limit the generalizability of their findings. Furthermore, our study included all clinical oro- and hypopharyngeal cancer patients treated with IMPT during the study period, making them a representative group of patients. Additionally, the study was conducted in a fully blinded manner, minimizing the likelihood of bias influencing the results. The clear preference for automated IMPT plans observed in our study provides further support for the clinical adoption of automated planning tools for IMPT. A limitation, however, is that the current workflow is not yet integrated with a clinical treatment planning system, which hampers direct clinical implementation.

Interobserver variability is a well-recognized challenge in radiotherapy, particularly due to subjective differences among ROs when assessing plan quality.[30] In this study, additional qualitative feedback was gathered through optional free-text comments. Despite individualized observations, all ROs preferred the automated plans over their manually generated counterparts in nearly all cases. This demonstrates that a single, standardized wish-list can produce clinically acceptable and, in most instances, preferred automated IMPT plans for HNC. These findings highlight the potential of automated planning not only to enhance planning efficiency but also to improve consistency in clinical practice.

In a clinical setting, physicians may have limited capacity due to high workloads and patient care demands. As a result, subtle details in treatment plans that are not immediately apparent may be overlooked, potentially affecting clinical acceptability or plan preference. In this retrospective study, however, there were no time constraints during plan assessment, and the ROs in this study could complete the questionnaires in between their tasks and at a time that suited them, ensuring time limitations could not have influenced the outcomes.

The automatically IMPT plans were generated within a few hours, comparable to the approximately 6-hour duration of manual clinical IMPT planning.[17] When clinical plan adaptations are needed due to anatomical changes such as tumor shrinkage or weight loss, our automated planning method is highly suitable for

generating high-quality, offline adapted plans, without requiring hands-on planning time. However, as the process generally takes hours, for online adaptations such as daily anatomical position changes or sinus filling, our method is currently not fast enough. Improving the speed of this is therefore an area of future research.

In a recent blinded comparison study involving deep learning IMPT plans for oropharyngeal cancer, additional manual refinements were required before achieving clinically acceptable results, while still not preferred always over the manual plans.[23] In contrast, the automatically generated IMPT plans in our study were not subjected to any manual fine-tuning, yet were preferred over the manually created plans. Our findings thus indicate not only an improvement in overall IMPT plan quality for HNC, but also potential for substantial time savings.

CONCLUSION

All fully automated IMPT plans for HNC were considered clinically acceptable by experienced HNC radiation oncologists. Automated plans were unanimously preferred over manually created plans in nearly all cases, with statistical significance, highlighting the clinical potential of automated IMPT planning tools.

REFERENCES

1. Nelms BE, Robinson G, Markham J, Velasco K, Boyd S, Narayan S, et al. Variation in external beam treatment plan quality: An inter-institutional study of planners and planning systems. Pract Radiat Oncol 2012;2:296–305. https://doi.org/10.1016/J.PRRO.2011.11.012.
2. 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.
3. Nystrom H, Jensen MF, Nystrom PW. Treatment planning for proton therapy: what is needed in the next 10 years? Br J Radiol 2020;93:20190304. https://doi.org/10.1259/BJR.20190304.
4. 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.
5. 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.
6. 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.
7. Hussein M, South CP, Barry MA, Adams EJ, Jordan TJ, Stewart AJ, et al. Clinical validation and benchmarking of knowledge-based IMRT and VMAT treatment planning in pelvic anatomy. Radiother Oncol 2016;120:473–9. https://doi.org/10.1016/J.RADONC.2016.06.022,.
8. 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;62:5926–44. https://doi.org/10.1088/1361-6560/AA71F8.
9. 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. https://doi.org/10.1186/S13014-015-0533-2.
10. 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. https://doi.org/10.1016/J.CTRO.2016.08.001.
11. 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.
12. 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.

13. 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.

14. Hedrick SG, Petro S, Ward A, Morris B. Validation of automated complex head and neck treatment planning with pencil beam scanning proton therapy. J Appl Clin Med Phys 2022;23:e13510. https://doi.org/10.1002/ACM2.13510;WGROUP:STRING:PUBLICATION.

15. 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 Biol Phys 2023;115:1283–90. https://doi.org/10.1016/J.IJROBP.2022.12.004.

16. Borderias-Villarroel E, Huet Dastarac M, Barragán-Montero AM, Helander R, Holmstrom M, Geets X, et al. Machine learning-based automatic proton therapy planning: Impact of post-processing and dose-mimicking in plan robustness. Med Phys 2023;50:4480–90. https://doi.org/10.1002/MP.16408.

17. Huiskes M, Kong W, Oud M, Crama K, Rasch C, Breedveld S, et al. Validation of Fully Automated Robust Multicriterial Treatment Planning for Head and Neck Cancer IMPT. Int J Radiat Oncol Biol Phys 2024;119:968–77. https://doi.org/10.1016/J.IJROBP.2023.12.034.

18. Vandewinckele L, Claessens M, Dinkla A, Brouwer C, Crijns W, Verellen D, et al. Overview of artificial intelligence-based applications in radiotherapy: Recommendations for implementation and quality assurance. Radiother Oncol 2020;153:55–66. https://doi.org/10.1016/J.RADONC.2020.09.008.

19. 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 2013;85:866–72. https://doi.org/10.1016/J.IJROBP.2012.04.015.

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. Fiandra C, Zara S, Richetto V, Rossi L, Leonardi MC, Ferrari P, et al. Multi-centre real-world validation of automated treatment planning for breast radiotherapy. Phys Medica 2024;123:103394. https://doi.org/10.1016/J.EJMP.2024.103394.

22. Vanderstraeten B, Goddeeris B, Vandecasteele K, van Eijkeren M, De Wagter C, Lievens Y. Automated Instead of Manual Treatment Planning? A Plan Comparison Based on Dose-Volume Statistics and Clinical Preference. Int J Radiat Oncol 2018;102:443–50. https://doi.org/10.1016/J.IJROBP.2018.05.063.

23. van Bruggen IG, van Dijk M, Brinkman-Akker MJ, Löfman F, Langendijk JA, Both S, et al. Clinical implementation of deep learning robust IMPT planning in oropharyngeal cancer patients: A blinded clinical study. Radiother Oncol 2024;200:110522. https://doi.org/10.1016/J.RADONC.2024.110522.

24. 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.

25. 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.

26. Kong W, Oud M, Habraken SJM, Huiskes M, Astreinidou E, Rasch CRN, et al. SISS-MCO: large scale sparsity-induced spot selection for fast and fully-automated robust multi-criteria optimisation of proton plans. Phys Med Biol 2024;69. https://doi.org/10.1088/1361-6560/AD1E7A.

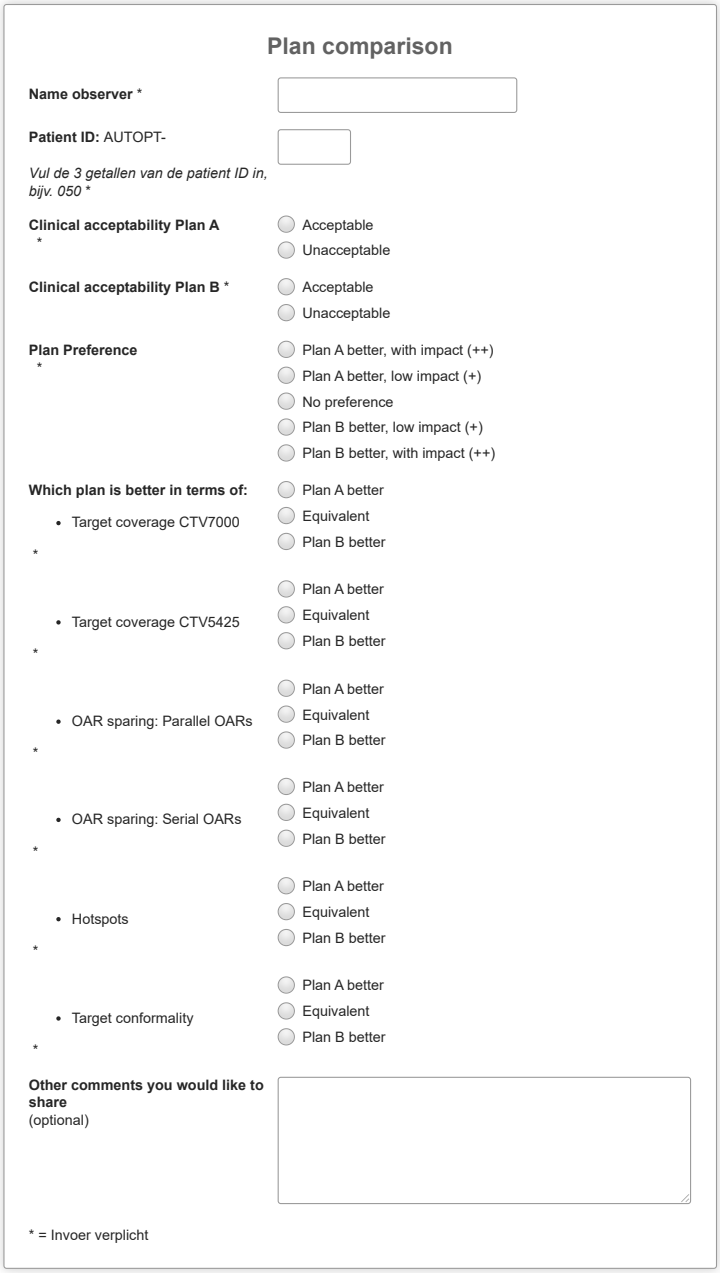
27. Kooy HM, Clasie BM, Lu HM, Madden TM, Bentefour H, Depauw N, et al. A case study in proton pencil-beam scanning delivery. Int J Radiat Oncol Biol Phys 2010;76:624–30. https://doi.org/10.1016/J.IJROBP.2009.06.065.

28. 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.

29. 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.

30. Cagni E, Botti A, Rossi L, Iotti C, Iori M, Cozzi S, et al. Variations in Head and Neck Treatment Plan Quality Assessment Among Radiation Oncologists and Medical Physicists in a Single Radiotherapy Department. Article 2021;11:1. https://doi.org/10.3389/fonc.2021.706034.

APPENDIX



Plan comparison

Name observer *

Patient ID: AUTOPT-

Vul de 3 getallen van de patient ID in, bijv. 050 *

Clinical acceptability Plan A *
- Acceptable
- Unacceptable

Clinical acceptability Plan B *
- Acceptable
- Unacceptable

Plan Preference *
- Plan A better, with impact (++)
- Plan A better, low impact (+)
- No preference
- Plan B better, low impact (+)
- Plan B better, with impact (++)

Which plan is better in terms of:

- Target coverage CTV7000 *
  - Plan A better
  - Equivalent
  - Plan B better
- Target coverage CTV5425 *
  - Plan A better
  - Equivalent
  - Plan B better
- OAR sparing: Parallel OARs *
  - Plan A better
  - Equivalent
  - Plan B better
- OAR sparing: Serial OARs *
  - Plan A better
  - Equivalent
  - Plan B better
- Hotspots *
  - Plan A better
  - Equivalent
  - Plan B better
- Target conformality *
  - Plan A better
  - Equivalent
  - Plan B better

Other comments you would like to share
(optional)

* = Invoer verplicht

Figure A1: The questionnaire used in this study

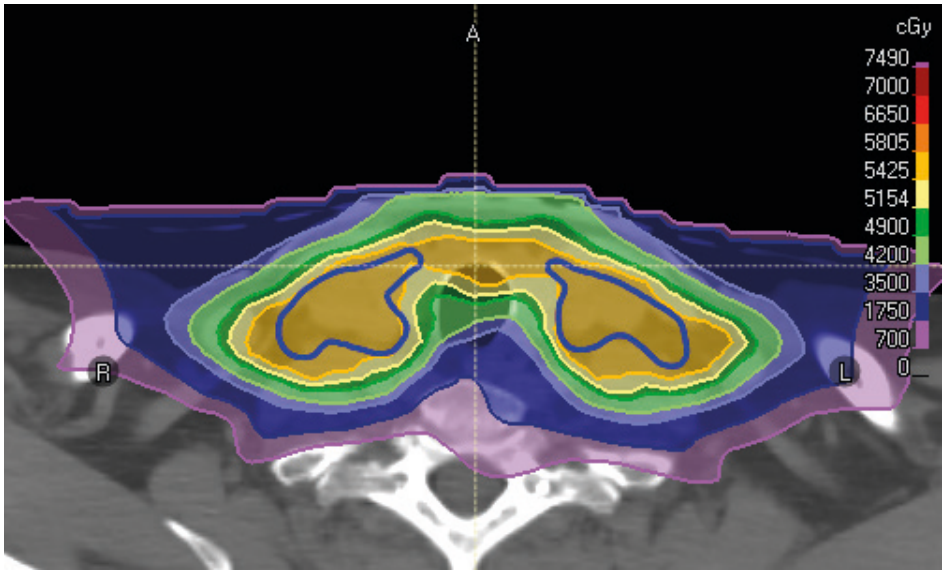


Figure A2: Example of an axial slice with dose distribution of a clinical plan which was considered clinically unacceptable, due to relative high dose located between the left and right parts of CTV5425 (delineated in blue).

Table A1: Preferences per patient and per RO

Patient	RO1	RO2	RO3
1	Auto plan better, low impact (+)	Auto plan better, with impact (++)	Auto plan better, with impact (++)
2	Auto plan better, low impact (+)	Auto plan better, low impact (+)	Auto plan better, low impact (+)
3	Auto plan better, low impact (+)	Auto plan better, low impact (+)	Auto plan better, with impact (++)
4	Auto plan better, low impact (+)	Auto plan better, low impact (+)	Auto plan better, low impact (+)
5	Auto plan better, low impact (+)	Auto plan better, low impact (+)	Auto plan better, with impact (++)
6	Auto plan better, with impact (++)	Auto plan better, with impact (++)	Auto plan better, with impact (++)
7	Auto plan better, low impact (+)	Auto plan better, with impact (++)	Auto plan better, with impact (++)
8	Auto plan better, low impact (+)	Auto plan better, with impact (++)	Auto plan better, low impact (+)
9	Auto plan better, with impact (++)	Auto plan better, with impact (++)	Auto plan better, low impact (+)
10	Auto plan better, low impact (+)	No preference	Auto plan better, with impact (++)
11	Auto plan better, low impact (+)	Auto plan better, low impact (+)	Auto plan better, with impact (++)
12	Auto plan better, low impact (+)	Auto plan better, low impact (+)	Auto plan better, low impact (+)
13	Auto plan better, low impact (+)	Auto plan better, low impact (+)	Auto plan better, low impact (+)
14	Auto plan better, with impact (++)	Auto plan better, low impact (+)	Auto plan better, with impact (++)
15	Auto plan better, low impact (+)	Auto plan better, with impact (++)	Auto plan better, with impact (++)
16	Auto plan better, low impact (+)	Auto plan better, with impact (++)	Auto plan better, low impact (+)
17	Auto plan better, low impact (+)	Auto plan better, low impact (+)	Auto plan better, low impact (+)
18	Auto plan better, low impact (+)	Auto plan better, low impact (+)	Auto plan better, low impact (+)
19	Auto plan better, with impact (++)	Auto plan better, with impact (++)	Auto plan better, with impact (++)
20	Auto plan better, low impact (+)	Auto plan better, low impact (+)	Auto plan better, with impact (++)

21	Auto plan better, with impact (++)	Auto plan better, with impact (++)	Auto plan better, low impact (+)
22	Auto plan better, low impact (+)	Auto plan better, low impact (+)	Auto plan better, with impact (++)
23	No preference	Auto plan better, with impact (++)	Auto plan better, low impact (+)
24	Auto plan better, low impact (+)	Auto plan better, low impact (+)	Auto plan better, with impact (++)
25	Auto plan better, with impact (++)	Auto plan better, with impact (++)	Auto plan better, with impact (++)
26	Auto plan better, with impact (++)	Auto plan better, low impact (+)	Auto plan better, low impact (+)
27	Auto plan better, with impact (++)	Auto plan better, low impact (+)	Auto plan better, with impact (++)
28	Auto plan better, with impact (++)	Auto plan better, low impact (+)	Auto plan better, low impact (+)
29	Auto plan better, low impact (+)	Auto plan better, low impact (+)	Auto plan better, with impact (++)
30	Auto plan better, with impact (++)	Auto plan better, low impact (+)	Auto plan better, low impact (+)