

## Scientific Article

# Development and Clinical Implementation of an Automated Virtual Integrative Planner for Radiation Therapy of Head and Neck Cancer



Elizabeth M. Jaworski, MD, MS, Michelle L. Mierzwa, MD, Karen A. Vineberg, MS, John Yao, PhD, Jennifer L. Shah, MD, Caitlin A. Schonewolf, MD, MS, Dale Litzenberg, PhD, Laila A. Gharzai, MD, LLM, Martha M. Matuszak, PhD, Kelly C. Paradis, PhD, Ashley Dougherty, CMD, Pamela Burger, CMD, Daniel Tatro, CMD, George Spencer Arnould, CMD, Jean M. Moran, PhD, Choonik Lee, PhD, Avraham Eisbruch, MD, and Charles S. Mayo, PhD\*

Department of Radiation Oncology, University of Michigan, Ann Arbor, Michigan

Received December 21, 2021; accepted July 10, 2022

## Abstract

**Purpose:** Head and neck (HN) radiation (RT) treatment planning is complex and resource intensive. Deviations and inconsistent plan quality significantly affect clinical outcomes. We sought to develop a novel automated virtual integrative (AVI) knowledge-based planning application to reduce planning time, increase consistency, and improve baseline quality.

**Methods and Materials:** An in-house write-enabled script was developed from a library of 668 previously treated HN RT plans. Prospective hazard analysis was performed, and mitigation strategies were implemented before clinical release. The AVI-planner software was retrospectively validated in a cohort of 52 recent HN cases. A physician panel evaluated planning limitations during initial deployment, and feedback was enacted via software refinements. A final second set of plans was generated and evaluated. Kolmogorov-Smirnov test in addition to generalized evaluation metric and weighted experience score were used to compare normal tissue sparing between final AVI planner versus respective clinically treated and historically accepted plans. A *t* test was used to compare the interactive time, complexity, and monitor units for AVI planner versus manual optimization.

**Results:** Initially, 86% of plans were acceptable to treat, with 10% minor and 4% major revisions or rejection recommended. Variability was noted in plan quality among HN subsites, with high initial quality for oropharynx and oral cavity plans. Plans needing revisions were comprised of sinonasal, nasopharynx, P-16 negative squamous cell carcinoma unknown primary, or cutaneous primary sites. Normal tissue sparing varied within subsites, but AVI planner significantly lowered mean larynx dose (median, 18.5 vs 19.7 Gy; *P* < .01) compared with clinical plans. AVI planner significantly reduced interactive optimization time (mean, 2 vs 85 minutes; *P* < .01).

**Conclusions:** AVI planner reliably generated clinically acceptable RT plans for oral cavity, salivary, oropharynx, larynx, and hypopharynx cancers. Physician-driven iterative learning processes resulted in favorable evolution in HN RT plan quality with significant time savings and improved consistency using AVI planner.

© 2022 The Authors. Published by Elsevier Inc. on behalf of American Society for Radiation Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Sources of support: Funding provided by Varian Medical Systems.  
Disclosures: Grant funding from Varian Medical Systems (J. M. M. and C. S. M.).

Data sharing statement: Research data are stored in an institutional repository and will be shared upon request to the corresponding author.

\*Corresponding author; E-mail: [cmayo@med.umich.edu](mailto:cmayo@med.umich.edu)

## Introduction

Radiation therapy (RT) is a cornerstone of head and neck (HN) cancer treatment. Intensity modulated radiation

<https://doi.org/10.1016/j.adro.2022.101029>

2452-1094/© 2022 The Authors. Published by Elsevier Inc. on behalf of American Society for Radiation Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

therapy (IMRT) has improved treatment accuracy and reduced RT-associated morbidity.<sup>1-10</sup> HN IMRT manual optimization is resource-intensive and variable, with heavy reliance upon physician and facility expertise.<sup>11-16</sup> HN IMRT implementation has been met with frequent treatment planning and quality assurance deviations, which are associated with worse outcomes.<sup>17-19</sup> Furthermore, the time required for HN IMRT planning must be considered in the context of survival advantages associated with minimizing total treatment time and time interval between consultation and starting treatment.<sup>20,21</sup> HN RT delivered at high-accruing centers is associated with improved outcomes, though factors including travel burden and patients' resources influence access to these centers.<sup>22-24</sup>

Automated planning has been developed to standardize treatment planning, maximize efficiency, improve plan quality, and mitigate geographic disparities by increasing access to high-quality RT plans.<sup>13,25</sup> Knowledge-based planning (KBP) models rely upon dosimetric and geometric experience from dose-volume histograms (DVH) of previously treated acceptable plans.<sup>25</sup> KBP benefits have been documented in various disease sites, including HN.<sup>26-34</sup> Iterative learning, a process incorporating manually driven feedback into model training, improves automated HN plan quality.<sup>35</sup> However, commercially available KBP algorithms are limited by smaller training data sets, lack of standardized inputs, challenging user-interface for plan revision, and limited ability to customize commercial algorithms to fit specific clinical needs. Script-based approaches like ours enable clinic-specific customization. Prior studies have characterized plan quality in cohorts of HN patients without regard for primary site, while others report achievements in only 1 subsite (eg, oropharynx<sup>36</sup> or nasopharynx<sup>31</sup>). There is a paucity of data regarding automated planning algorithm performance among different HN sites.

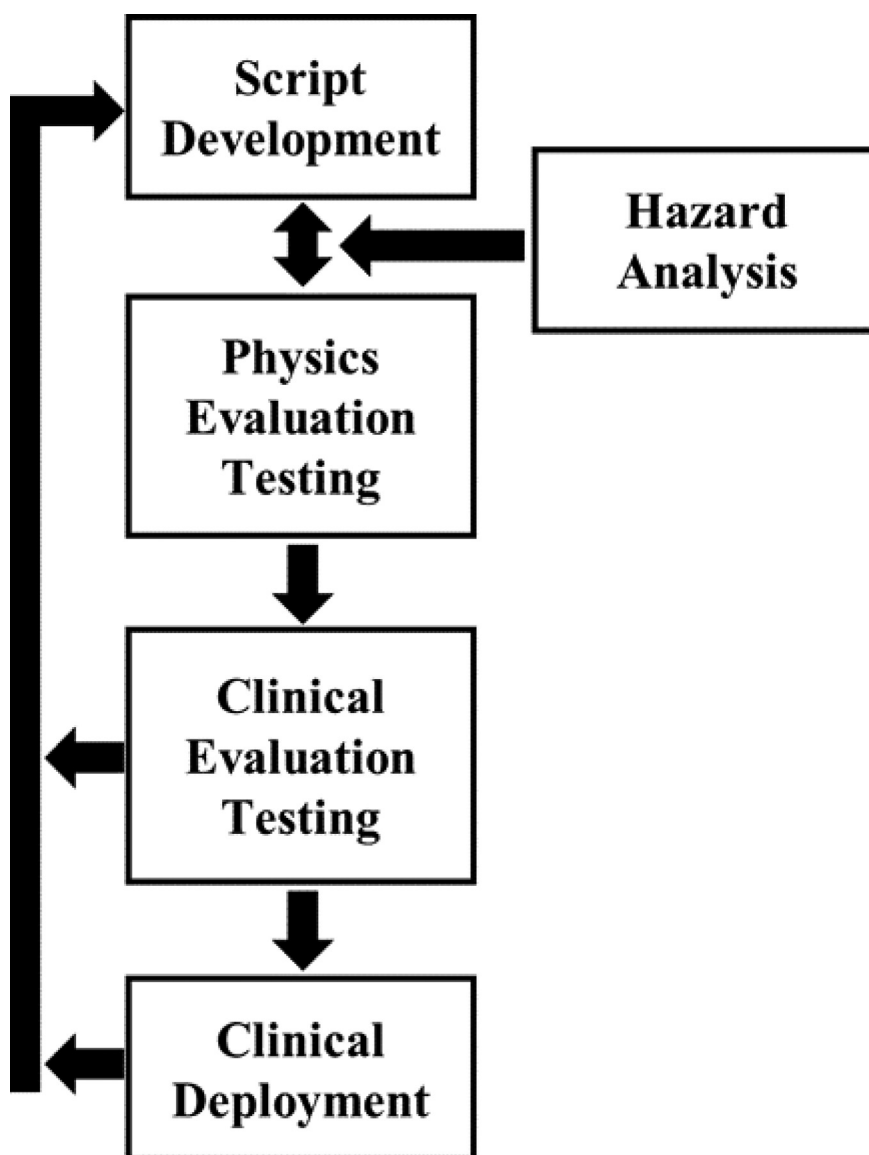
Herein, we report the development of an automated virtual integrative (AVI) planning algorithm. The algorithm is not a machine learning approach. This algorithm was designed using the same treatment planning system tools applied by dosimetrists during the manual process and integrates historical optimization norms from prior plans. The AVI-planner algorithm uniquely generates optimization parameters based upon statistical analyses of DVH metrics from previously treated HN RT plans. We sought to create preliminary automated HN RT plans for "warm start optimization," where dosimetrists continue optimization from the automated plan instead of starting each plan with a new manual process.<sup>37</sup> We describe the iterative learning process to address planning deficiencies noted for select primary sites. To our knowledge, this is the first investigation of an HN-specific automated planning algorithm whereby the identification of site-specific clinically significant deficiencies drives autoplanner script refinements to improve overall RT plan quality.

## Methods and Materials

### Script development and hazard analysis

Our script release process is shown in [Figure 1](#). The write-enabled script was developed to incorporate practice norms defined by a library of 668 previously treated HN RT plans collected at our institution between 2014 and 2019. This library was comprised of 31.3% oropharynx (n = 209), 19.3% oral cavity (n = 129), 14.7% larynx (n = 98), 7.9% cutaneous (n = 53), 6.6% salivary (n = 44), 4.2% sinonasal (n = 28), 3.7% nasopharynx (n = 25), 3% unknown primary (n = 20), 2.7% hypopharynx (n = 18), 1.8% thyroid (n = 12), 0.6% orbital or lacrimal (n = 4), and 4.2% "other" (n = 28). Software inputs were standardized including nomenclature and complete sets of contoured organs at risk/planning target volumes (OARs/PTVs) with explicitly defined planning priorities and objectives. Within the foundational library, >90% of plans contained spinal cord, brain stem, bilateral cochlea, parotids, superior and inferior pharyngeal constrictors, oral cavity, esophagus, mandible, and lips. When surgically present and clinically relevant, bilateral submandibular glands (SMGs) were included in 75%, larynx in 81%, bilateral optic nerves, chiasm, eyes, and lenses were included in 18% to 25%, while only 11% included lacrimal glands (data not shown).

During development, the AVI-planner algorithm statistically evaluated DVH parameters from the 668 plan library, which then informed optimization parameters. Optimization constraints were defined as less than 30% of historic values. A team of physicists, dosimetrists, and software developers then used the algorithm to iteratively optimize a subset of 20 HN patients. None of the 20 HN plans were included in the physician round 1 evaluation. Before round 1 evaluation (see following sections), all planning parameters in the algorithm were finalized for physician evaluation. Based upon standardized input targets and OARs, the algorithm created a full set of optimization structures using typical margin and Boolean operations. Optimization structures included subvolumes of overlapping OARs and target structures, as well as high-dose PTV subvolumes segmented from lower PTV volumes. Dose sculpting rings were used by the normal tissue objective to conform prescription isodose lines. The AVI-planner software automatically placed an isocenter, segmented optimization structures, and generated beams and plan setup with full calculation. All plans were volumetric-modulated arc therapy, calculated in Eclipse version 15.6, with the analytical anisotropic algorithm, using 0.25-cm grid size. Eclipse Scripting Application Programming Interface (ESAPI) enabled the integration of AVI-planner software with Eclipse (Varian Medical Systems, Palo Alto, CA). The nonclinical, research version of ESAPI mimicked manual optimization and allowed



**Fig. 1** Flow chart depicting write-enabled script development and release process.

interaction with the optimizer during optimization. However, the clinical ESAPI version did not allow this interaction. Because our objective was designing software compatible with the Food and Drug Administration approved ESAPI versions, our interface and algorithm generated HN plans that could be sequentially, manually modified after optimization. Optimization with our AVI-planner algorithm did not allow for dynamic real-time interaction with the optimizer.

Routine physics quality plan check was employed for the automated plans, which then proceeded on to a second phase of clinical evaluation. Before clinical use, a prospective hazard analysis was performed using a streamlined failure mode and effects analysis described by Paradis et al.<sup>38</sup> A process map for clinical use of the script was generated with associated hazards (failure modes) from multidisciplinary feedback. The priority score for

each failure mode (a version of the relative risk priority number from TG-100) was assigned as high, medium, or low.<sup>39</sup> All failure modes with high or medium priority scores were mitigated before proceeding to plan evaluation and clinical deployment.

### Patient selection

This study was institutional review board exempt (HUM 00126332) for quality improvement. AVI planner in round 1 optimization was retrospectively validated within a cohort of 52 patients with HN cancer treated between 2019 and 2020. None of these 52 plans were included within the foundational 668 plan library. We included oral cavity, oropharynx, larynx, hypopharynx, cutaneous, sinonasal, and salivary primaries to account

for anatomy and OARs, adjuvant versus definitive RT, target dose, and fractionation. Institutional dose-escalation and de-escalation protocol patients were included. We excluded hypofractionated and palliative patients. Simulation computed tomography scans were performed on a Philips Brilliance big-bore 16 slice scanner (Koninklijke Philips N.V., Amsterdam, Netherlands) using 3-mm slices. Patients were scanned head-first, supine with intravenous contrast, and immobilized in 5-point thermoplastic masks. Intact and postoperative boost and elective clinical target volume contours were delineated referencing published guidelines<sup>40,41</sup> with a 3-mm PTV margin. Dosimetrists manually optimized clinical plans using Eclipse (Varian Medical Systems, Palo Alto, CA), which were delivered on Varian TrueBeam or Clinac linear accelerators with 120 leaf multileaf collimator using 6-MV photons with 2 to 4 volumetric-modulated arc therapy arcs.

## Plan evaluation

Clinical plans underwent peer review by a subspecialty panel of attending radiation oncologists. Institutional protocols specified prioritization of target coverage and objectives for OAR sparing (Table E1). To identify AVI-planner limitations consistently requiring additional manual input for warm start optimization, the physician panel evaluated clinical acceptability of round 1 AVI-planner cases. HN subsites were grouped by treatment paradigm and anatomic proximity. These plans were rejected if the plan was unsafe and unsalvageable despite reoptimization. Major revisions indicated a high perceived risk of either (1) a clinically relevant toxicity due to exceeded OAR constraints or (2) risk of recurrence from target undercoverage. Plans with minor revisions were safe with room for improvement in conformality, heterogeneity, or target coverage. The highest quality plans were deemed “treat as is.” Physician feedback from round 1 was addressed per “write-enabled script refinement.” All 52 cases were then replanned with the AVI-planner script without manual modifications and labeled “round 2.” The same physician panel re-evaluated all 52 plans.

Beyond stand-alone clinical acceptability, round 2 AVI-planner quality was compared with (1) clinically treated plans, (2) historically accepted plans, and (3) literature-based thresholds.<sup>42</sup> “Clinically treated plan” denotes the patient-specific RT plan, which was delivered during the patient’s treatment course. Within this context, evaluating round 2 versus the clinically treated plan provides an individual, patient-level comparison of plan quality. Comparisons to historically accepted plans were based on summarized metrics captured from the entire 668 HN foundational library. Thus, round 2 plan quality was assessed in the context of aggregate institutional experience with all 668 considered high-quality HN plans.

Evaluating round 2 plans in both situations more fully characterizes plan quality at both the patient level and institutional experience level.

To compare AVI planner to historic plans, constraint metrics within the algorithm were derived from 668 previously treated plans using the previously described generalized evaluation metric (GEM) and weighted experience score (WES) described by Mayo et al.<sup>43</sup> GEM compares DVH metrics to constraints and historical values, which are cast onto a sigmoidal curve with a scale of 0 to 1, where  $GEM = 0.5$  if the constraint was met and 0.95 when 95% of historical values were lower than the current plan’s value. WES ranks the DVH curves with respect to historical values, on a 0 to 1 scale, with values weighted according to historic variability. WES correlates with normal tissue complication probability but rises sooner with respect to dose, correlating with physician preferences to drive doses below normal tissue complication probability thresholds.

VRxGy[%] was used to assess coverage at the prescribed dose for each dose level. The International Commission on Radiation Units and Measurements (ICRU) conformality index (CI)<sup>26,44</sup> was calculated for PTV\_High, PTV\_Low, and PTV\_Mid00 volume:

$$CI_{ICRU} = \frac{Body : VRx[cc]}{PTV : Volume[cc]}$$

Dose heterogeneity within PTVs was assessed using ICRU 83 HI.<sup>45</sup>

$$HI_1 = \frac{(D2\%[Gy] - D98\%[Gy])}{D50\%[Gy]}$$

These were calculated for the PTV subvolumes, not overlapping with volumes at prescribed doses as PTV\_High, PTV!\_Low, and PTV!\_Mid00 in TG-263 nomenclature.

We collected total monitor units (MU) per plan for the 52 patient cohort as well as calculated complexity described by Younge et al<sup>46</sup>:

$$M = \frac{1}{MU} \sum_{i=1}^N MU_i \times \frac{y_i}{A_i}$$

## Write-enabled script refinement and clinical deployment

Iterative learning occurred by a 2-step process, which used physician feedback to refine the optimization algorithm. The first iteration of plans reflected the explicitly stated prescription planning objectives using statistical data gained from the 668 HN plan library (round 1). In round 1, a template of optimization constraints was defined by quantile analysis of DVH metrics within our plan library of 668 previously treated patients. For

structures evaluated by mean (Gy), the constraint corresponded with the lower 30% of historic values for D90% (Gy), D50% (Gy), and D10% (Gy). This enabled prioritization of portions receiving lower dose further away from PTV (D90% [Gy]) compared with portions receiving high dose in close proximity to PTVs (D10% [Gy]). Round 1 planning used the validation cohort described in the following sections, and “round 1” indicates optimization with the initially released AVI-planner script and minor manual edits. After round 1 evaluation, we discovered additional implicit physician preferences and expectations that were not stated in the prescription planning documentation. The algorithm was modified in several ways to incorporate physician feedback. This modified, refined algorithm was subsequently used to generate refined plans (round 2). To shorten development time required to refine algorithm performance, these modified parameters were placed in an external configuration file. This limited the scope of changes that required recompiling the code and also facilitates more rapid customization in the future when releasing this script to other clinics. Script modifications were as follows:

1. Normal tissue constraints and priorities. Instead of limiting the level of priority to 1, 2, or 3, we included additional granular priority levels (ie, priority 1, 1.5, 2, 2.5, 3, etc) to better align with physician intent. Instead of a fixed constraint value, the algorithm was modified to allow increasing or decreasing a given constraint.
2. Dose-sculpting structures. Automatically generated rings and buffers were added to increase conformality and minimize dose in nontarget and non-OAR normal tissues (ie, minimize low and intermediate dose within the base of the neck).
3. Segmented structures. Subvolumes of PTV and OAR overlap were transformed into standardized segmented structures, and new constraints and priorities were added to enhance the algorithm’s ability to more precisely control dose. For example, areas of parotid and PTV overlap were segmented out, and a higher priority for sparing was placed upon the nonoverlapping ipsilateral parotid.
4. Isocenter placement. In response to major revision or rejected plans, we modified the algorithm’s ability to detect unilateral or atypical target location and, subsequently, adapt isocenter placement was added. This maximized use of the central 0.5 cm multileaf collimator leaves.

After the evaluation process described, AVI planner was deployed to the clinic as a staged process. The initial use of the application was carefully monitored by physician and physicist stakeholders in the initial limited clinical release for proper functioning and introduction of hazards. After validation, the script was then deployed

without changes. The script is routinely used in clinic, though dosimetrists regularly manually modify these automated plans with physician input. Requests for additional improvements and features are monitored and incorporated into future development cycles.

## Statistical analysis

We used Python 3.8 statistical software for this analysis. A 1-sided, Kolmogorov-Smirnov test was used to determine whether the distribution of AVI-planner OAR mean or D0.1cc values was higher or lower than clinically treated plans of the validation cohort. The distribution of per plan differences was analyzed. AVI-planner values were compared with literature-based thresholds<sup>42</sup> using a normal distribution, with matched cardinality, centered on each threshold with a 0.5 Gy standard deviation as the reference distribution using a *t* test for mean value difference. Descriptive statistics were used to evaluate dosimetric endpoints and time required for treatment planning. One-tailed *t* test was used to compare the interactive time required for AVI planning versus manual planning. Two-tailed *t* test was used for comparing total MUs and complexity between AVI planner and clinical plans.

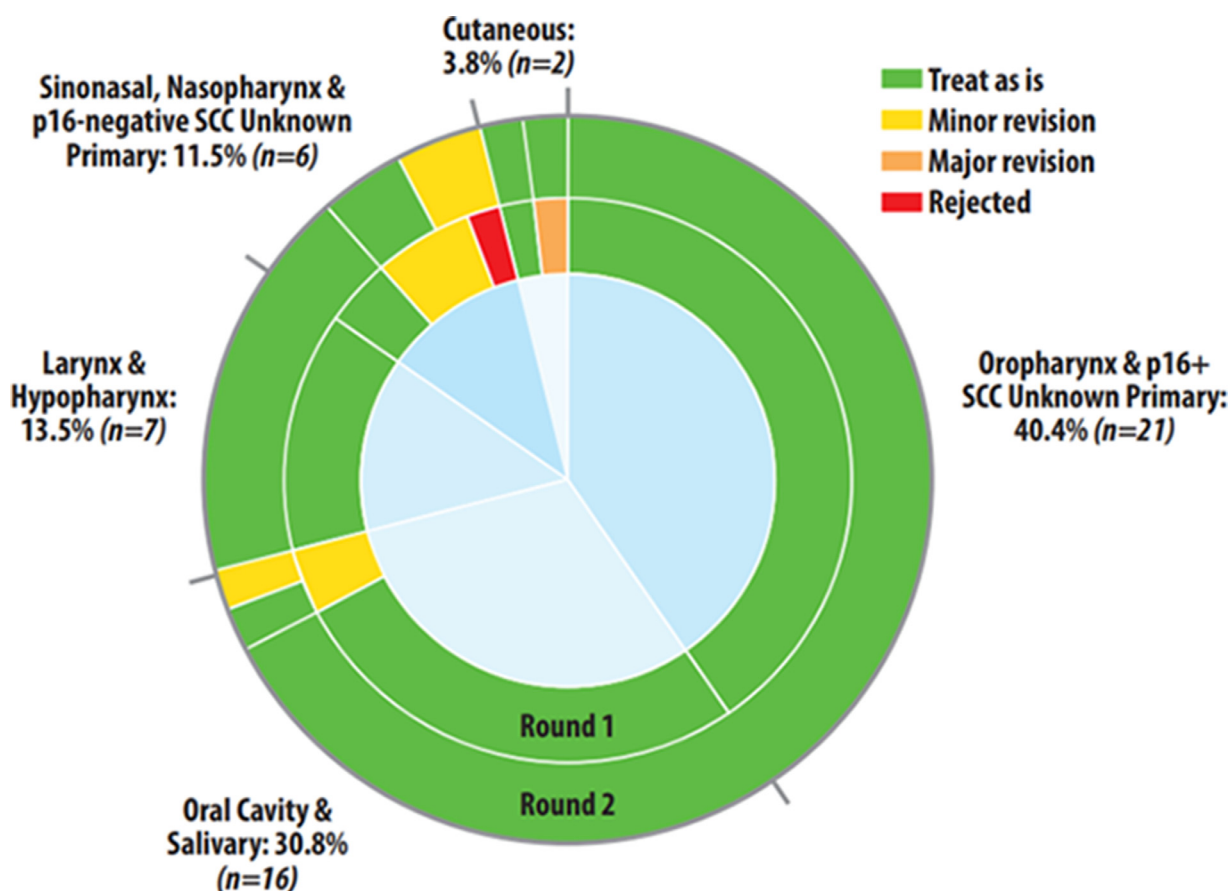
## Results

### Failure mode and effects analysis

Before clinical deployment, 12 failure modes were identified relating to contour generation (7), plan creation (1), treatment field generation (2), plan optimization (1), and plan approval (1) (Table E2). None of these failure modes were higher relative risk compared with the manual treatment planning process. A detailed summary of the failure modes and associated mitigations is shown in Table E2. Several code modifications were prompted by failure mode and effects analysis. These included detailed analysis of structure volumes at the beginning and end of the algorithm to identify changes made; checks at entry of the script algorithm that PTV and OAR volumes are approved and cannot be edited; and enforcement of naming conventions for structure sets, course, and plans to minimize risk of unintentional use of an automated plan.

### Validation and clinical implementation: Round 1

We retrospectively validated AVI planner in 52 patients, which consisted of mostly men (69%) with locally advanced oropharynx (40%) or oral cavity and salivary (31%) cancers (Fig 2, Table E3). Definitive intent



**Fig. 2** Highlighting the importance of disease subsite-specific automated virtual integrative-planner algorithm performance. We show an integrated pie and doughnut chart demonstrating clinical acceptability of automated virtual integrative-planned cases among head and neck subsites. Central chart (blue) shows plan frequency by subsite ( $n = 52$ ). Innermost doughnut chart shows “round 1” clinical acceptability. Outermost doughnut chart, “round 2,” shows evolution in acceptability for 7 plans initially requiring revisions.

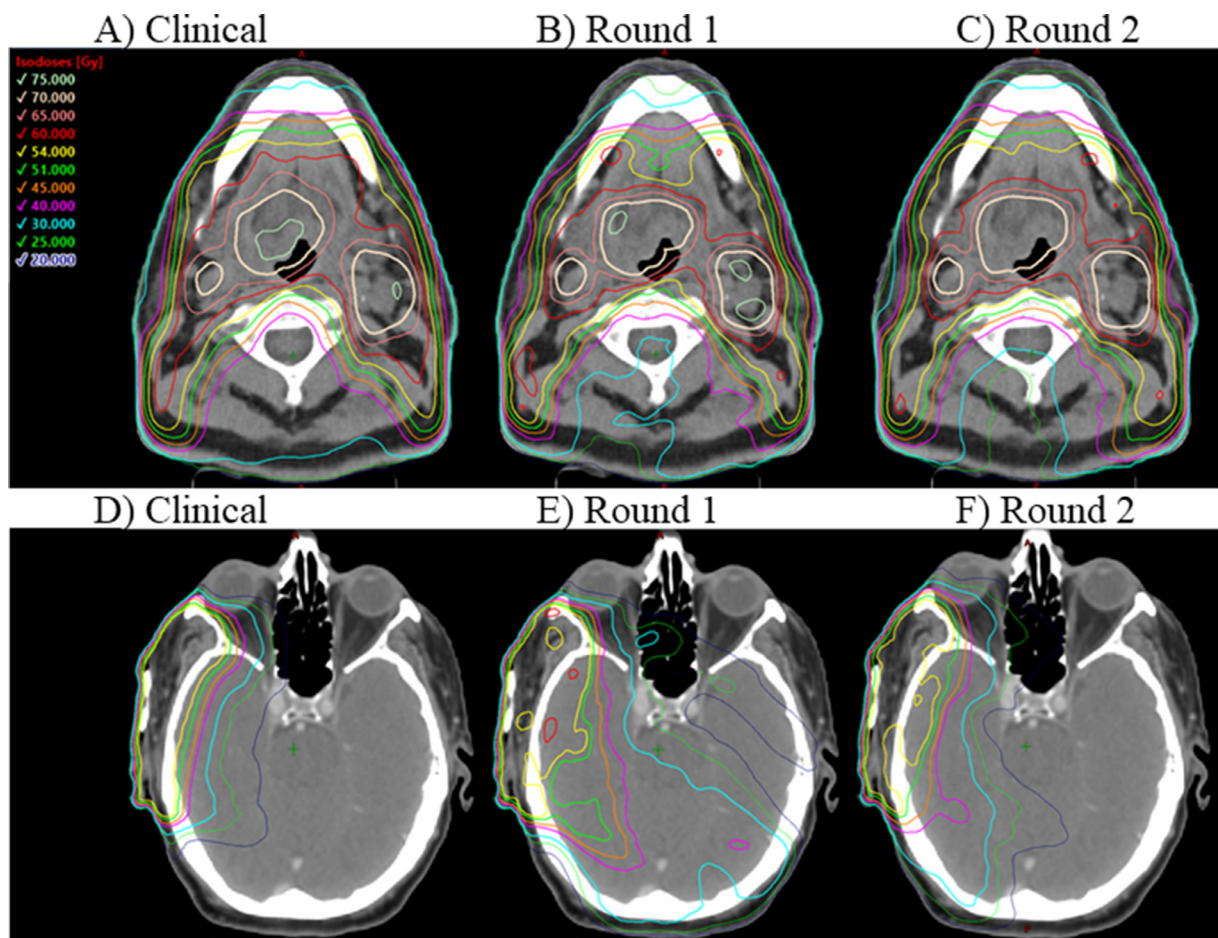
organ-preservation RT comprised the majority of plans (58%) with a median of 70 Gy (range, 54–80 Gy) in 35 fractions (range, 27–35 fx); 62% received concurrent chemotherapy.

Overall, 86% of round 1 plans were safe to treat; however, we identified variability in plan quality among different HN subsites (Fig 2). All oropharynx and p16+ squamous cell carcinoma (SCC) unknown primary (21/21 plans) and larynx and hypopharynx (7/7 plans) were “treat as is.” Similarly, most oral cavity and salivary cases (14/16 plans; 87.5%) required no revisions. This contrasts with the frequency of major revisions or rejections recommended for sinonasal, nasopharynx, and p16-negative SCC unknown primary (1/6 plans; 17%) and cutaneous (1/2 plans; 50%; Fig 2). Minor revisions were increasing conformality and reducing heterogeneity. Major revisions were limiting hot spots outside PTV, restricting hot spots within PTV to 105% to 110%, and improving target coverage. Sample round 1 isodose distributions are shown for both a definitive early stage p16+ base of tongue cancer considered “treat as is” (Fig 3B) compared with “major

revisions” for an adjuvantly treated malar cheek Merkel cell carcinoma (Fig 3E). Software refinements were made to the script in response to the round 1 evaluation, which included normal tissue constraints and priorities, dose-sculpting structures, segmented structures, and isocenter placement, as discussed in the section “Write-enabled script refinement and clinical deployment.”

### Clinical reassessment: Round 2

During round 2 evaluation of all 52 plans, there were no rejections or major revisions (Fig 2). Minor revisions were recommended for 1 oral cavity (6.3%) and 3 sinonasal or nasopharynx or p16-negative SCC unknown primary plans (50%). The remaining 48 plans were “treat as is.” Minor revisions in round 2 focused on improving conformality or more aggressive sparing of spinal cord, optics, and contralateral orbit or salivary structures. This evolution in quality is evident for the adjuvantly treated Merkel cell carcinoma (Fig 3D–F).



**Fig. 3** Identifying the strengths and limitations of automated planning among typical versus atypical head and neck cases. We show sample plans for early-stage oropharynx and adjuvant Merkel cell carcinoma. Top panel shows a stage II cT2N2M0 p16+ squamous cell carcinoma of the right base of tongue treated with definitive chemoradiation to 70 Gy in 35 fx A-C, which is considered a typical case and well-represented within the model. The bottom panel shows a stage III pT1 pN1a(sn) Merkel cell carcinoma treated adjuvantly to 54 Gy in 30 fx D-F, which is considered atypical and underrepresented within the model. Left panels A,D, show clinically treated plan. Middle panels show B, round 1 automated virtual integrative-planner “treat as is”; E, round 1 AVI-planner “major revision” due to conformality and 119% hotspot outside planning target volume. Right panels show C,F, round 2 after automated virtual integrative-planner upgrades. Isodose lines (absolute dose, Gy) show 75 (light green), 70 (white), 65 (pink), 60 (red), 54 (yellow), 51 (green), 45 (orange), 40 (purple), 30 (cyan), 25 (green), and 20 (dark blue).

Conformality and dose heterogeneity for all PTV levels were similar between round 2 AVI-planner and clinical cases (Table 1). The number of MUs per plan was significantly lower for AVI-planner round 2 (mean,  $619.7 \pm 69.7$  MU) compared with the clinically treated plan ( $693.5 \pm 219.4$  MU;  $P = .03$ ). Similarly, AVI planner generated less complex plans compared with clinically treated plans (mean complexity score,  $0.13 \pm 0.02$  vs  $0.14 \pm 0.03$ ;  $P < .01$ ). To evaluate patterns of OAR sparing in round 2 AVI planner, we compared the entire distribution of mean or D0.1cc dose values between AVI-planner cases versus clinically treated or historically accepted plan values (Fig 4A). Given that oropharynx was the most prevalent HN subsite within both the foundational library and the validation cohort, we also selected a representative DVH

from a locally advanced oropharynx cancer treated with definitive chemoradiation. This demonstrates typical DVH metrics from a case that was well represented in the model (Fig 5).

Considering all 52 plans, the contralateral parotid dose distribution was higher with AVI planner compared with clinically treated plans (median, 25 vs 23 Gy;  $P < .01$ ), with higher doses compared with historic plans (WES, 0.50 vs 0.42;  $P < .01$ ). Inferior pharyngeal constrictor muscles had higher distribution of mean dose in AVI planner versus clinical plans (median, 21 vs 19 Gy;  $P < .01$ ), with higher doses compared with historic plans (WES, 0.53 vs 0.35;  $P < .01$ ), and narrowly exceeded our constraint (GEM 0.52). Conversely, AVI planner lowered the dose to ipsilateral SMG (62 vs 65 Gy;  $P = .04$ ), though

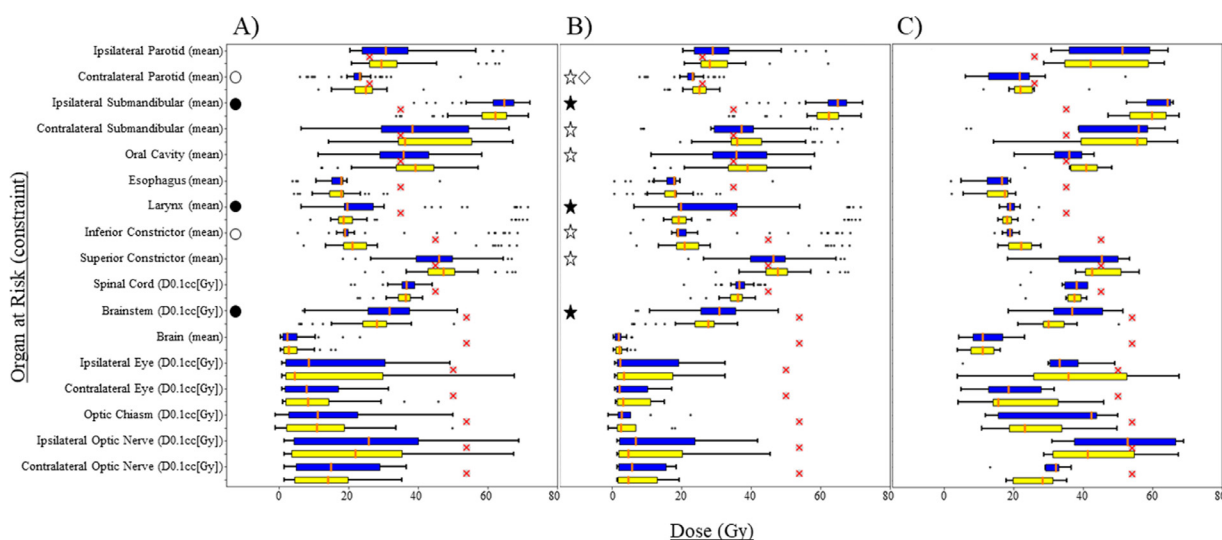
**Table 1** Conformality and heterogeneity (ICRU 83) indexes of clinical and round 2 AVI planner for high, intermediate, and low PTV

	PTV_High			PTV_Mid			PTV_Low		
	Clinical plan	AVI planner	P value	Clinical plan	AVI planner	P value	Clinical plan	AVI planner	P value
Conformality index	1.1 ± 0.7	1.2 ± 0.7	.7	1.3 ± 0.4	1.4 ± 0.5	.7	1.3 ± 0.2	1.3 ± 0.2	.8
Heterogeneity index	1.1 ± 0.04	1.1 ± 0.03	.5	1.1 ± 0.04	1.1 ± 0.04	.4	1.2 ± 0.1	1.2 ± 0.1	.4

Abbreviations: AVI = automated virtual integrative; ICRU = international commission on radiation units and measurements; PTV = planning target volume.

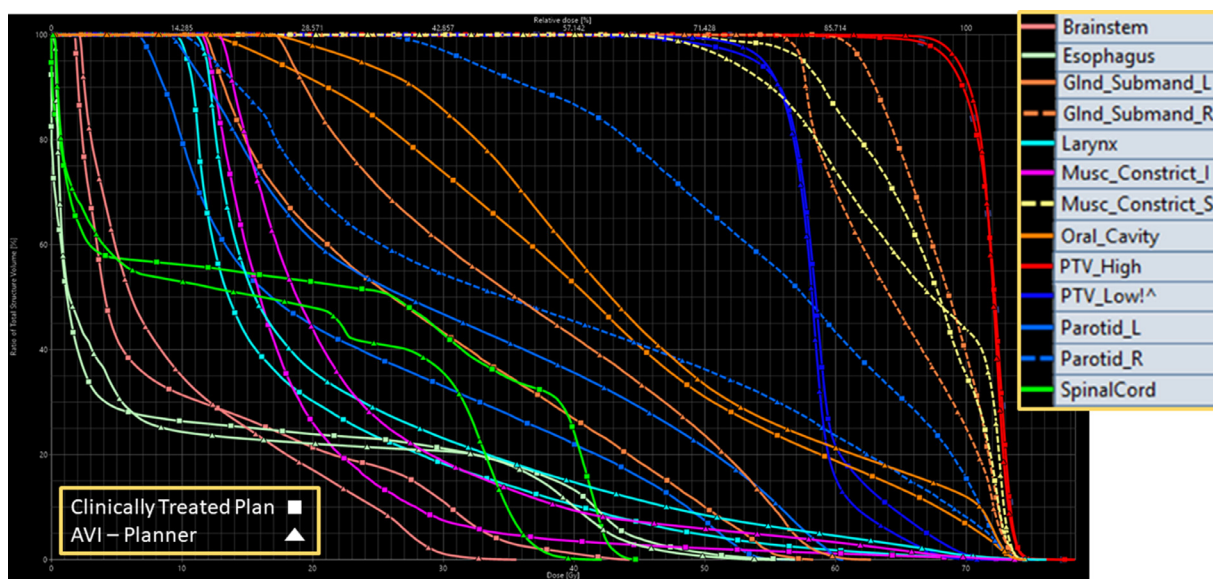
this was not clinically relevant (GEM >0.90).<sup>43</sup> AVI planner lowered the distribution of mean dose to the larynx compared with clinical (median, 19 vs 20 Gy;  $P < .01$ ) and historical plans (WES, 0.29 vs 0.44;  $P < .01$ ). Brain stem D0.1cc from AVI planner was lower than that in clinical plans (median, 28 vs 32 Gy;  $P < .01$ ). Spinal cord distribution of D0.1cc was lower in AVI planner compared with historic plans (WES, 0.44 vs 0.63;  $P < .01$ ), but similar to clinical plans (median, 36.4 vs 36.7 Gy;  $P = .9$ ). Distribution of dose to optic nerves, chiasm, eyes, contralateral SMG, superior pharyngeal constrictors, oral cavity, mandible, and esophagus were similar among AVI-planner cases, clinical, and historic plans (Fig 4A).

For oropharynx or p16+ SCC unknown primary (n = 21; Fig 4B), the distribution of mean larynx dose was significantly lower with AVI planner versus clinical plans (median, 18 vs 20 Gy;  $P < .01$ ) or historical plans (WES, 0.28 vs 0.44;  $P < .01$ ), which was clinically relevant (GEM 0.46). Esophagus and ipsilateral parotid were spared equally among AVI planner, clinical, and historic plans. Distribution of mean dose to contralateral SMG was higher for AVI planner compared with clinical plans (median, 36 vs 33 Gy;  $P = .02$ ) and historical plans (WES, 0.44 vs 0.41;  $P = .02$ ). Neither clinical nor AVI planner met constraints for relevant sparing (GEM, 0.62 and 0.57). Contralateral parotid distribution of mean dose



**Fig. 4** Differential normal tissue sparing by clinical versus automated virtual integrative (AVI)-planner. Comparisons for all head and neck subsites A; oropharynx, p16+ squamous cell carcinoma unknown primary, oral cavity, salivary, larynx, hypopharynx B; sinonasal, nasopharynx, and p16-negative SCC unknown primary C. Organs at risk are listed on y-axis with corresponding constraint. Dose (Gy) on x-axis. Box plots of clinical (blue) or round 2 AVI-planner (yellow) provide median, interquartile range, minimum, and maximum doses. Red “x” denotes consensus thresholds.<sup>42</sup> The difference in normal tissue sparing between AVI-planner and clinical plan is typically small compared with the difference in relation to established thresholds. Statistical significance was achieved with  $P < .05$  on 1-sided Kolmogorov-Smirnov test. Symbols along y-axis indicate statistically significant difference in organs at risk sparing between AVI-planner versus clinical plans: total cohort (panel A, circles), oropharynx and p16+ squamous cell carcinoma unknown primary (panel B, stars), oral cavity and salivary (panel B, diamonds). Filled shapes indicate AVI-planner significantly improved sparing whereas unshaded symbols indicate clinical plan achieved significantly better sparing.





**Fig. 5** Representative typical dose-volume histogram for a cT4N1M0 p16+ squamous cell carcinoma of the left tonsil treated definitively to 70 Gy, comparing manual plan (squares) versus automated virtual integrative-planner (triangles). X-axis is dose (Gy), Y-axis volume (%).

was higher for AVI planner compared with clinical plans (median, 25 vs 23 Gy;  $P = .047$ ), but similar to historically accepted plans (WES, 0.50 vs 0.43;  $P = .1$ ). Superior and inferior pharyngeal constrictors received higher dose with AVI planner and exceeded constraints compared with historical controls ( $P < .01$ ). Oral cavity distribution of mean dose was higher with AVI planner versus clinical (37 vs 33 Gy;  $P = .04$ ) and historic plans (WES, 0.64 vs 0.53;  $P = .04$ ; Fig 4B).

OAR doses were similar between clinical and round 2 AVI planner for the remaining HN subsites (Fig 4B, C). Of note, the oral cavity/salivary contralateral parotid distribution of mean dose was significantly higher for AVI planner compared with clinical plans (median, 26 vs 23 Gy;  $P < .01$ ) and historic plans (WES, 0.53 vs 0.42;  $P < .01$ ) and did not meet constraints (GEM 0.56) (Fig 4B). Two cutaneous plans did not reach the 3 plan threshold required for formal comparison.

### Time study

We compared Eclipse optimizer interactive time for 10 recent manual plans versus interactive time with AVI-planner software for 51 of the validation cohort patients. This interactive time included all steps of the manual optimization, such as segmentation structures, setting isocenter, ring and buffer dose sculpting structures, normal tissue optimization limits, target and OAR prioritization, setting the number of arcs, and optimization time. Of the 10 manually optimized plans, 70% were oropharynx ( $n = 7$ ), while 30% were comprised of oral cavity ( $n = 1$ ),

thyroid ( $n = 1$ ), and unknown primary ( $n = 1$ ). Mean time for manual interaction time was shorter for AVI planner versus manually optimized plans, 2 versus 85 minutes, respectively ( $P < .01$ ).

### Discussion

We developed and implemented a knowledge-based AVI software to facilitate HN treatment planning. Initially, we identified inconsistent plan quality among different HN subsites. After iterative software adaptations, we noted favorable evolution in target coverage, heterogeneity, and OAR sparing. This software exceeded our warm start optimization goal and rapidly created clinically acceptable plans without manual adjustments for many HN subsites. We have published the source code for AVI planner at a GitHub repository (<https://github.com/CSMayoLab/AVIplanner.git>) for educational purposes only, to promote use and development by researchers of their own automated planning solutions.

Regarding clinical acceptability of automated HN plans, we found 86% of round 1 plans were treat as is, which is comparable to 88% by Radiation Planning Assistant.<sup>47</sup> Our script was developed from a diverse training data set, capturing unique nuances and planning considerations. The inconsistent site-specific plan quality likely resulted from limited experience within the foundational library. Our heterogeneous library accrued over 5 years, but there were fewer cutaneous (7.9%), sinonasal (4.2%), and nasopharynx (3.7%) plans. Improvements in both conformality and heterogeneity were shown for prostate and cervix cancer after refining Varian's RapidPlan

default settings,<sup>48</sup> but studies detailing specific software refinements and evolution in plan quality among multiple subsites are limited for HN.<sup>35,47</sup>

Physicians frequently emphasized higher OAR prioritization. For instance, the clinical plan aggressively spared contralateral parotid further below the planning objective in a cT1N1 p16+ tonsil cancer, whereas AVI planner less aggressively spared the contralateral parotid to meet constraints. Given OAR constraint heterogeneity, we compared the round 2 AVI planner results to consensus thresholds.<sup>42</sup> Structure laterality is reported; however, the distinction of ipsilateral or contralateral relative to the target is less readily available. AVI planner achieved lower contralateral parotid doses (median, 25 Gy) than the 26 Gy threshold ( $P < .01$ ). AVI planner lowered dose to optic structures, eyes, brain stem, spinal cord, esophagus, inferior constrictors, and mandible compared with accepted thresholds. Larynx doses achieved in AVI-planner cases were lower than thresholds (35 Gy;  $P < .01$ ) and values with automation approaches by Fogliata et al<sup>27</sup> ( $24.8 \pm 5.2$  Gy) or Ouyang et al<sup>26</sup> (28.4 Gy). This highlights the importance of benchmarking automation achievements against literature values, historic norms, and the validation subset.

Secondarily, the failure modes addressed during development can be found in manual planning, suggesting these hazards already exist and may be more likely to happen without the software. Thus, automated planning does not obviate standard clinical and physics quality assurance. Similar to Wang et al,<sup>49</sup> inconsistencies in standardized OAR prioritization affected the performance of our model. In line with time-savings noted by other groups for autoplanned nasopharynx<sup>31,50,51</sup> and oropharynx cancers,<sup>33,36</sup> we confirmed time savings with AVI planner. AVI planner generated less complex plans ( $P < .01$ ) with fewer MUs ( $P = .03$ ) compared with the clinically treated plans.

Limitations of this work include that this software is integrated only with Eclipse for 30 to 35 fraction plans. Given the revisions required for sinonasal, nasopharynx, and cutaneous sites, this software should be used cautiously near the skull base. Three target dose levels are currently supported, but additional dose levels require manual editing. Dosimetrists must also ensure the relevance of automated decisions. For instance, planners must remain vigilant about modifying the isocenter location or number of arcs for a unilateral target. Physicians must explicitly address planning preferences in the planning directive. For example, in a locally advanced maxillary sinus cancer requiring adjuvant RT after an orbital exenteration, aggressively sparing the remaining contralateral orbit and lacrimal gland may take precedence over PTV coverage. The user-friendly interface contains the same tools used in manual optimization, allowing real-time modification by dosimetrists, compared with fully automated optimization, which must run to completion before permitting revision. However, these standardized optimization parameters likely differ from personalized approaches of experienced

dosimetrists. Therefore, additional time may be required for revisions. Interinstitutional heterogeneity in delineated OARs, inconsistent OAR contouring, and variability in constraints and prioritization are barriers to widespread adoption of automated planning.

To our knowledge, this is the first report identifying HN primary site-specific variability in automated plan quality, which favorably evolved with physician input. Our work cautions against interpreting that automated planning achievements are universal among HN subsites. This is relevant for clinics that would ideally employ 1 planning algorithm for all HN cases, instead of separate optimization algorithms for each HN subsite. We are not advocating this software in lieu of skilled dosimetrists, clinical expertise or patients seeking choice of treatment at experienced centers of care. The use cases outlined demonstrate the potential of the solution and the positive benefits in settings with limited resources, increased demand, urgent starts or reduced subspecialized dosimetrists. The AVI-planner software and associated integration into workflows demonstrate the ability to create and scale the availability of high quality HN RT plans.

Furthermore, our institutional adoption of AVI planner into routine practice has expanded the number of dosimetrists able to rapidly generate high quality HN plans. We plan to release AVI planner to our affiliate sites to improve plan quality and uniformity, and we are also extending this standardized approach to other disease sites including lung and prostate. In the future, automated planning will facilitate adaptive RT planning. Future software upgrades may incorporate generalized equivalent uniform dose, update the foundational library, and focus on well-lateralized cases near the skin surface. Application programming interfaces that enable clinics to programmatically automate all parts of the treatment planning process give clinics the tools they need to increase efficiency and consistency in plan quality in their process workflows. To promote these clinical improvements, it is highly desirable for manufacturers to provide application programming interfaces that give users, at minimum, the same capabilities they have in manual operations to algorithmically interact with optimizers, dose calculation engines, reference points, and scheduling capabilities.

## Acknowledgments

The authors would like to acknowledge Steven Kronenberg for his assistance with the creation of the figures for this manuscript.

## Supplementary materials

Supplementary material associated with this article can be found in the online version at [doi:10.1016/j.adro.2022.101029](https://doi.org/10.1016/j.adro.2022.101029).

## References

- Hawkins PG, Kadam AS, Jackson WC, Eisbruch A. Organ-sparing in radiotherapy for head-and-neck cancer: Improving quality of life. *Semin Radiat Oncol*. 2018;28:46–52.
- Nutting CM, Morden JP, Harrington KJ, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): A phase 3 multicentre randomised controlled trial. *Lancet Oncol*. 2011;12:127–136.
- Pow EH, Kwong DLW, McMilan AD, et al. Xerostomia and quality of life after intensity-modulated radiotherapy vs. conventional radiotherapy for early-stage nasopharyngeal carcinoma: Initial report on a randomized controlled clinical trial. *Int J Radiat Oncol Biol Phys*. 2006;66:981–991.
- Kam MK, Leung S-F, Zee B, et al. Prospective randomized study of intensity-modulated radiotherapy on salivary gland function in early-stage nasopharyngeal carcinoma patients. *J Clin Oncol*. 2007;25:4873–4879.
- Dirix P, Nuyts S. Evidence-based organ-sparing radiotherapy in head and neck cancer. *Lancet Oncol*. 2010;11:85–91.
- Lee N, Puri DR, Blanco AI, Chao KSC. Intensity-modulated radiation therapy in head and neck cancers: An update. *Head Neck*. 2007;29:387–400.
- Eisbruch A, Ship JA, Martel MK, et al. Parotid gland sparing in patients undergoing bilateral head and neck irradiation: Techniques and early results. *Int J Radiat Oncol Biol Phys*. 1996;36:469–480.
- Murdoch-Kinch CA, Kim HM, Vineberge KA, Ship JA, Eisbruch A. Dose-effect relationships for the submandibular salivary glands and implications for their sparing by intensity modulated radiotherapy. *Int J Radiat Oncol Biol Phys*. 2008;72:373–382.
- Feng FY, Kim HM, Lyden TH, et al. Intensity-modulated chemoradiotherapy aiming to reduce dysphagia in patients with oropharyngeal cancer: Clinical and functional results. *J Clin Oncol*. 2010;28:2732–2738.
- Beadle BM, Liao K-P, Elting LS, et al. Improved survival using intensity-modulated radiation therapy in head and neck cancers: A SEER-Medicare analysis. *Cancer*. 2014;120:702–710.
- Boero IJ, Paravati AJ, Xu B, et al. Importance of radiation oncologist experience among patients with head-and-neck cancer treated with intensity-modulated radiation therapy. *J Clin Oncol*. 2016;34:684–690.
- Lee CC, Huang T-T, Lee M-S, et al. Survival rate in nasopharyngeal carcinoma improved by high caseload volume: A nationwide population-based study in Taiwan. *Radiat Oncol*. 2011;6:92.
- Cilla S, Ianiro A, Romano C, et al. Template-based automation of treatment planning in advanced radiotherapy: A comprehensive dosimetric and clinical evaluation. *Sci Rep*. 2020;10:423.
- Batumalai V, Jameson MG, Forstner DF, Vial P, Holloway LC. How important is dosimetrist experience for intensity modulated radiation therapy? A comparative analysis of a head and neck case. *Pract Radiat Oncol*. 2013;3:e99–e106.
- Moore KL, Brame RS, Low DA, Mutic S. Experience-based quality control of clinical intensity-modulated radiotherapy planning. *Int J Radiat Oncol Biol Phys*. 2011;81:545–551.
- Nelms BE, Robinson G, Markham J, 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.
- Eisbruch A, Harris J, Garden AS, et al. Multi-institutional trial of accelerated hypofractionated intensity-modulated radiation therapy for early-stage oropharyngeal cancer (RTOG 00-22). *Int J Radiat Oncol Biol Phys*. 2010;76:1333–1338.
- Zhong H, Men K, Wang J, et al. The impact of clinical trial quality assurance on outcome in head and neck radiotherapy treatment. *Front Oncol*. 2019;9:792.
- Peters LJ, O'Sullivan B, Giralt J, et al. Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: Results from TROG 02.02. *J Clin Oncol*. 2010;28:2996–3001.
- Graboyes EM, Kompelli AR, Neskey DM, et al. 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.
- Rosenthal DI, Liu L, Lee JH, et al. Importance of the treatment package time in surgery and postoperative radiation therapy for squamous carcinoma of the head and neck. *Head Neck*. 2002;24:115–126.
- Wuthrick EJ, Zhang Q, Machtay M, et al. Institutional clinical trial accrual volume and survival of patients with head and neck cancer. *J Clin Oncol*. 2015;33:156–164.
- Naghavi AO, Echevarri MI, Strom TJ, et al. Patient choice for high-volume center radiation impacts head and neck cancer outcome. *Cancer Med*. 2018;7:4964–4979.
- George JR, Yom SS, Wang SJ. Combined modality treatment outcomes for head and neck cancer: Comparison of postoperative radiation therapy at academic vs nonacademic medical centers. *JAMA Otolaryngol Head Neck Surg*. 2013;139:1118–1126.
- Hussein M, Heijmen BJM, Verellen D, Nisbet A. Automation in intensity modulated radiotherapy treatment planning—a review of recent innovations. *Br J Radiol*. 2018;91: 20180270.
- Ouyang Z, Liu Shen Z, Murray E, et al. Evaluation of auto-planning in IMRT and VMAT for head and neck cancer. *J Appl Clin Med Phys*. 2019;20:39–47.
- Fogliata A, Reggiori G, Stravato A, et al. RapidPlan head and neck model: The objectives and possible clinical benefit. *Radiat Oncol*. 2017;12:73.
- Krayenbuehl J, Norton I, Studer G, Guckenberger M. Evaluation of an automated knowledge based treatment planning system for head and neck. *Radiat Oncol*. 2015;10:226.
- Tol JP, Delaney AR, Dabele M, Slotman BJ, Verbakel WF. Evaluation of a knowledge-based planning solution for head and neck cancer. *Int J Radiat Oncol Biol Phys*. 2015;91:612–620.
- Gintz D, Latifi K, Caudell J, et al. Initial evaluation of automated treatment planning software. *J Appl Clin Med Phys*. 2016;17:331–346.
- Giaddui T, Geng H, Chen Q, et al. Offline quality assurance for intensity modulated radiation therapy treatment plans for NRG-HN001 head and neck clinical trial using knowledge-based planning. *Adv Radiat Oncol*. 2020;5:1342–1349.
- Hansen CR, Bertelsen A, Hazell I, et al. Automatic treatment planning improves the clinical quality of head and neck cancer treatment plans. *Clin Transl Radiat Oncol*. 2016;1:2–8.
- Kusters JMAM, Bzdusek K, Kumar P, et al. Automated IMRT planning in Pinnacle: A study in head-and-neck cancer. *Strahlenther Onkol*. 2017;193:1031–1038.
- Krayenbuehl J, Zamburlini M, Ghandour S, et al. Planning comparison of five automated treatment planning solutions for locally advanced head and neck cancer. *Radiat Oncol*. 2018;13:170.
- Fogliata A, Cozzi L, Reggiori G, et al. RapidPlan knowledge based planning: Iterative learning process and model ability to steer planning strategies. *Radiat Oncol*. 2019;14:187.
- Kamima T, Ueda Y, Fukunaga J-I, et al. Multi-institutional evaluation of knowledge-based planning performance of volumetric modulated arc therapy (VMAT) for head and neck cancer. *Phys Med*. 2019;64:174–181.
- Ahunbay EE, Ates O, Li XA. An online replanning method using warm start optimization and aperture morphing for flattening-filter-free beams. *Med Phys*. 2016;43:4575.
- Paradis KC, Naheedy KW, Matuszak MM, Kashani R, Burger P, Moran JM. The fusion of incident learning and failure mode and effects analysis for data-driven patient safety improvements. *Pract Radiat Oncol*. 2021;11:e106–e113.
- Huq MS, Fraass BA, Dunscombe PB, et al. The report of Task Group 100 of the AAPM: Application of risk analysis methods to radiation therapy quality management. *Med Phys*. 2016;43:4209.
- Biau J, Lapeyre M, Troussier I, et al. Selection of lymph node target volumes for definitive head and neck radiation therapy: A 2019 update. *Radiat Oncol*. 2019;134:1–9.

41. Gregoire V, Evans M, Le Q-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.
42. Lee AW, Ng WT, Pan JJ, et al. International guideline on dose prioritization and acceptance criteria in radiation therapy planning for nasopharyngeal carcinoma. *Int J Radiat Oncol Biol Phys.* 2019;105:567–580.
43. Mayo CS, Yao J, Eisbruch A, et al. Incorporating big data into treatment plan evaluation: Development of statistical DVH metrics and visualization dashboards. *Adv Radiat Oncol.* 2017;2:503–514.
44. Baltas D, Kolotas C, Geramani K, et al. A conformal index (COIN) to evaluate implant quality and dose specification in brachytherapy. *Int J Radiat Oncol Biol Phys.* 1998;40:515–524.
45. ICRU report 83: Prescribing, recording, and reporting photon-beam intensity-modulated radiation therapy (IMRT). *J ICRU.* 2010;10:35–36.
46. Younge KC, Roberts D, Janes LA, et al. Predicting deliverability of volumetric-modulated arc therapy (VMAT) plans using aperture complexity analysis. *J Appl Clin Med Phys.* 2016;17:124–131.
47. Olanrewaju A, Court LE, Zhange L, et al. Clinical acceptability of automated radiation treatment planning for head and neck cancer using the radiation planning assistant. *Pract Radiat Oncol.* 2021;11:177–184.
48. Hussein M, South CP, Barry MA, et al. Clinical validation and benchmarking of knowledge-based IMRT and VMAT treatment planning in pelvic anatomy. *Radiother Oncol.* 2016;120:473–479.
49. Wang Y, Heijmen BJM, Petit SF. Knowledge-based dose prediction models for head and neck cancer are strongly affected by interorgan dependency and dataset inconsistency. *Med Phys.* 2019;46:934–943.
50. Chang ATY, Hung AWM, Cheung FWK, et al. Comparison of planning quality and efficiency between conventional and knowledge-based algorithms in nasopharyngeal cancer patients using intensity modulated radiation therapy. *Int J Radiat Oncol Biol Phys.* 2016;95:981–990.
51. Hu J, Liu B, Xie W, et al. Quantitative comparison of knowledge-based and manual intensity modulated radiation therapy planning for nasopharyngeal carcinoma. *Front Oncol.* 2020;10:551763.