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 , Charles S. Mayo PhD

PII: S2452-1094(22)00135-X
DOI: https://doi.org/10.1016/j.adro.2022.101029
Reference: ADRO 101029

To appear in: *Advances in Radiation Oncology*

Received date: 21 December 2021
Accepted date: 10 July 2022

Please cite this article as: 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 , Charles S. Mayo PhD , Development and Clinical Implementation of an Automated Virtual Integrative Planner for Radiation Therapy of Head and Neck Cancer, *Advances in Radiation Oncology* (2022), doi: https://doi.org/10.1016/j.adro.2022.101029

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

Running title: "AVI-planner for head and neck cancer radiation"

Elizabeth M. Jaworski, MD, MS[1], Michelle L. Mierzwa, MD[1], Karen A. Vineberg, MS[1], John Yao, PhD[1], Jennifer L. Shah, MD[1], Caitlin A. Schonewolf, MD, MS[1], Dale Litzenberg, PhD, Laila A. Gharzai, MD, LLM[1], Martha M. Matuszak, PhD[1], Kelly C. Paradis, PhD[1], Ashley Dougherty, CMD[1], Pamela Burger, CMD[1], Daniel Tatro, CMD[1], George Spencer Arnould, CMD[1], Jean M. Moran, PhD[1], Choonik Lee, PhD[1], Avraham Eisbruch, MD[1], Charles S. Mayo, PhD[1]

[1]Department of Radiation Oncology, University of Michigan, Ann Arbor, Michigan

Corresponding Author:

Charles Mayo, PhD

University Hospital, B2C432

1500 East Medical Center Dr.

Ann Arbor, MI 48109-5010

cmayo@med.umich.edu

734-232-3837

Statistician: Charles S. Mayo, PhD cmayo@med.umich.edu

Funding statement: Funding provided by Varian Medical Systems.

Disclosure statement: Authors disclose the following: grant funding from Varian Medical Systems (JMM, CSM).

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

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

Abstract

Purpose: Head and neck (HN) radiation (RT) treatment planning is complex and resource intensive. Deviations and inconsistent plan quality significantly impact 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.

Materials and Methods: 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 (KS) test in addition to Generalized Evaluation Metric (GEM) and Weighted Experience Score (WES) were used to compare normal tissue sparing between final AVI-planner versus respective clinically treated and historically accepted plans. 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 SCC Unknown Primary or cutaneous primary sites. Normal tissue sparing varied within subsites, but AVI-planner significantly lowered mean larynx dose (median 18.5 Gy vs 19.7 Gy, $p<0.01$) compared to clinical plans. AVI-planner significantly reduced interactive optimization time (mean 2 vs 85 minutes, $p<0.01$).

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.

Introduction

Radiation therapy (RT) is a cornerstone of HN cancer treatment. Intensity-modulated radiation therapy (IMRT) has improved treatment accuracy and reduced RT-associated morbidity [1-10]. HN IMRT manual optimization is resource-intensive and variable, with heavy reliance upon physician and facility expertise [11-16]. HN IMRT implementation has been met with frequent treatment planning and quality assurance (QA) deviations, which are associated with worse outcomes [17-19]. 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 [20, 21]. 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 [22-24].

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 [13, 25]. Knowledge-based planning (KBP) models rely upon dosimetric and geometric experience from dose-volume histograms (DVH) of previously treated acceptable plans [25]. KBP benefits have been documented in various disease sites, including HN [26-34]. Iterative learning, a process incorporating manually driven feedback into model training, improves automated HN plan quality [35]. However, commercially available KBP algorithms are limited by smaller training datasets, 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 one subsite (e.g. oropharynx [36] or nasopharynx [31]). 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 [37]. We describe the iterative learning process to address planning deficiencies noted for select primary sites. To our knowledge, this is the first investigation of a HN-specific automated planning algorithm whereby the identification of site-specific clinically-significant deficiencies drive autoplanner script refinements to improve overall RT plan quality.

Methods:

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 2014-2019. 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), 4.2% "other" (n=28). Software inputs were standardized including nomenclature and complete sets of contoured organs at risk/planning target volumes (OAR/PTVs) with explicitly defined planning priorities and objectives. Within the foundational library, >90% of plans contained spinal cord, brainstem, bilateral cochlea, parotids, superior and inferior pharyngeal constrictors, oral cavity, esophagus, mandible, lips. When surgically present and clinically relevant, bilateral submandibular glands (SMG) were included in 75%, larynx in 81%, bilateral optic nerves, chiasm, eyes, and lenses were included in 18-25%, 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. Prior to Round 1 evaluation (see below), 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 sub-volumes of overlapping OAR 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 VMAT, calculated in Eclipse version 15.6, with the analytical anisotropic algorithm (AAA), using 0.25 cm grid size. Eclipse Scripting Application Programming Interface (ESAPI) enabled the integration of AVI-planner software with Eclipse (Varian Medical System, Palo Alto, CA). The non-clinical, research version of ESAPI mimicked manual optimization and allowed interaction with the optimizer during optimization. However, the clinical ESAPI version did not allow this interaction. Since our objective was designing software compatible with the Food and Drug Administration (FDA) approved ESAPI versions, our interface and algorithm generated HN plans which 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 onto 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. [38]. 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 [39]. All failure modes with high or medium priority scores were mitigated before proceeding to plan evaluation and clinical deployment.

Patient Selection

This study was IRB exempt (HUM 00126332) for quality improvement. AVI-planner in Round 1 optimization was retrospectively validated within a cohort of 52 HN cancer patients treated between 2019-2020. 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 vs definitive RT, target dose, and fractionation. Institutional dose-escalation or de-escalation protocol patients were included. We excluded hypofractionated and palliative patients. Simulation CT 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 IV contrast and immobilized in 5-point thermoplastic masks. Intact and postoperative boost and elective CTV contours were delineated referencing published guidelines [40, 41] with a 3 mm PTV margin. Dosimetrists manually optimized clinical plans using Eclipse (Varian Medical System, Palo Alto, CA), which were delivered on Varian TrueBeam or Clinac linear accelerators with 120 leaf MLC using 6-MV photons with 2-4 VMAT 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 (Supplementary Table 1). 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 under-coverage. 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 to 1) clinically treated plans 2) historically accepted plans and 3) literature-based thresholds [42]. 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. [43]. GEM compares DVH metrics to constraints and historical values, which are cast onto a sigmoidal curve with 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