# A Head and Neck Contour Grading System Provides an Objective Assessment of Radiation Oncology Resident Contouring Skills

Astha Rohit, BS; Diego A.S. Toesca, MD; Justin D. Gagneur, MA; Samir H. Patel, MD; Jean-Claude M. Rwigema, MD; Lisa McGee, MD*

## ABSTRACT

**Purpose:** This study evaluated the effectiveness of a peer-review process to objectively assess the skills of radiation oncology (RO) resident physicians (RPs) in contour grading of head and neck (HN) malignancies.

**Methods:** Target volumes from consecutive patients diagnosed with primary HN malignancies, treated in a single institution, were contoured by RPs during HN service rotations and were formally peer-reviewed by a minimum of 2 HN RO attendings and assigned a grade as follows: R0 (no change recommended); R1 (minor revision recommended, not clinically significant); and R2 (major revision recommended, deemed clinically significant). Progression of residents' HN contouring skills was assessed in accordance with their postgraduate year (PGY) in training.

**Results:** Formal contour peer review was performed for 218 patients with HN cancer contoured by 6 RO RPs from 2018 to 2024. Of those cases, 48 (22%) were contoured by PGY2 RPs, 98 (45%) by PGY3 RPs, 40 (18%) by PGY4 RPs, and 32 (15%) by PGY5 RPs. There was an objective improvement in contour grades and a reduced need for target volume modifications through the progression of academic years, with a mean score of 1.43 (SD = 0.71; CI = 0.2) for PGY2 trainees; 0.99 (SD = 0.81; CI = 0.16) for PGY3 trainees; 0.93 (SD = 0.92; CI = 0.29) for PGY4 trainees; and 0.69 (SD = 0.64; CI = 0.23) for PGY5 trainees. Improvement in scores was consistent among all RO RPs, with absolute mean improvements of −0.2 (RP #1), −0.32 (RP #2), −0.82 (RP #3), −0.4 (RP #4), −1.33 (RP #5), and −0.56 (RP#6).

**Conclusions:** Incorporating a formal HN contouring peer-review process and contour grade assignment into routine clinical evaluation of RO RPs provides an objective metric of their HN contour quality progression throughout training, by PGY. This tool can be used as an added, objective assessment of RO resident competency in contour evaluation.

**Keywords:** contouring peer review, target volume, target delineation, resident physician education, assessment, objective

## Introduction

The existing framework for monitoring the skill progression of radiation oncology (RO) resident physicians (RPs) throughout their postgraduate training has historically relied on subjective measures. Most institutions in the United States use the Accreditation Council for Graduate Medical Education (ACGME) milestones to evaluate RPs throughout their postgraduate training to ensure necessary competencies are met prior to graduation.[1,2] The ACGME Milestone 2.0 version outlines 6 key competencies:

Patient Care; Medical Knowledge, Practice-Based Learning and Improvement; Interpersonal and Communication Skills; Professionalism; and System-Based Practice. Within these competencies, there are subcompetencies, such as contouring and target delineation skill, that are graded on a scale from 1 to 5.[2] Every 6 months, programs' Clinical Competency Committees evaluate the RPs and determine their appropriate skill levels in all pertinent categories with the expectation that all RPs will reach level 4, at the minimum, in all sectors prior to their graduation.[2]

Due to the inherent subjectivity of the goals and the diversity of the field, the assessment of RPs' proficiency is largely left to the discretion of the evaluator.[1] Currently, the most overt objective measure appears to be the number of treatment cases performed or observed by the RP, with a minimum mandatory case attendance per disease site.[3] Despite recent advancements, there remains an imbalance of subjective and objective metrics, obscuring standardization and comparison among programs. Certain aspects of RO training are disproportionately affected, such as simulations, contouring, planning, treatment setup, and procedure proficiency, which are heavily skill-based and crucial to effective treatment.[1] These variations in training and monitoring practices make it difficult to gauge an RP's readiness for independent practice on a national scale.

In radiation therapy planning, the process of contouring involves delineating the target volume as well as avoidance structures and organs at risk, in order to maximize local tumor control and minimize treatment toxicity.[4] Contouring is a key skill for RPs to master as appropriate target volume delineation largely dictates the quality of a radiation therapy treatment plan and, ultimately, the oncological outcome for patients. Suboptimal contouring skills could therefore result in tumor recurrences or undesired side effects with the potential to significantly impact patient quality of life.[5-10]

Cancer of the head and neck (HN) is historically difficult to treat, exemplified by one of the highest inter-physician variabilities in contouring for the disease site.[5,6] A multitude of factors—including a complex local anatomy, high number of relatively small and sensitive anatomical structures in close proximity to one another, and anatomical variations between patients—can skew the standardization of contours in patients with HN cancer and contribute to one of the highest observed toxicities among cancer disease subsites commonly treated in RO.[11,12]

In an effort to improve patient outcomes by reducing interobserver variability, many institutions have implemented a peer-review process to optimize plan quality and patient safety.[9,13,14] Multiple professional organizations, including the American Society for Radiation Oncology (ASTRO), American College of Radiology, and Royal Australian and New Zealand College of Radiologists have recommended the use of a multiphysician team to review radiation therapy (RT) plans, provide feedback to treating physicians, and cultivate a safe environment that encourages performance improvement in areas such as contours and treatment planning.[9,13,15,16] These processes, and the aspects of RT planning that they evaluate, vary widely across institutions. ASTRO emphasizes the impact of contouring on patient outcomes, assigning the highest priority to target volume delineation in the peer-review process.[13] Repeatedly, institutional studies focused on contouring in HN cancers have shown that modifications resulting from peer review directly impact patient prognosis by reducing RT-induced toxicities and improving survival rates.[7,9,13,14,17]

The peer-review process fosters a collaborative, educational environment in which learning occurs through both participation and observation. It allows RPs to actively participate in treatment plan evaluations by asking questions and offering or receiving feedback, all of which can improve their confidence and skills while enhancing

their understanding and implementation of institutional guidelines.[9,14]

In this study, we report the outcomes of a prospective, formal, HN contour grading process that was developed during peer-review sessions at our institution. The protocol was used to objectively assess competency improvement in target volume delineation by RO RPs throughout their postgraduate training.

## Materials and Methods

### Study Design and Procedures

Following Institutional Review Board approval, a protocol was implemented to assess the quality of contouring completed by RO RPs for patients diagnosed with primary HN cancers. From 2018 to 2024, RO RPs on HN service used auto-contouring technology to establish target volumes for patients requiring RT treatment. HN contours were completed on a treatment planning CT scan, aided by the incorporation of information from diagnostic imaging (PET and/or MRI) fused to the treatment planning CT, as well as information from clinical examination (e.g., flexible laryngoscopy exam, operative notes and surgical pathology when appropriate). Unless clinically contraindicated, contrast was used for CT simulations.

Once their contours were completed, the RO RP cases were formally presented for peer review to a minimum of 2 RO HN attending physicians (APs) who provided appraisal consensus, grading, and feedback. Sessions were held weekly following HN tumor board review, with additional ad hoc sessions scheduled as needed. Target contours were reviewed by all available HN APs, with mandatory review by the treating HN oncologist and a minimum of one additional HN AP. During the peer-review process, contour edit recommendations and feedback were provided verbally to the RO RPs for each individual case. Initial contours were assigned a grade as follows: R0 (no change recommended); R1 (minor revision recommended, not clinically significant);

and R2 (major revision recommended, deemed clinically significant). An R1 grade reflected the need for stylistic changes to match the contours of the AP, rather than a decrement in RP skill set, while an R2 grade indicated shortcomings in the target contours that could negatively impact patient outcomes, such as the omitting gross residual disease or inaccurate coverage of the postoperative tumor bed. The grades were recorded in a peer-review task area of the patient's electronic medical record ARIA (Varian Medical Systems, Inc., Palo Alto, CA), alongside the initials of the resident and all peer-review physicians. After the recommended contouring edits were completed by the RO RP and re-reviewed by the AP responsible for the case, contours were sent to dosimetry to initiate treatment planning.

Because the contour grade assigned to each case was determined by a collective consensus among HN RO APs, inter-reviewer variability was not assessed.

### Statistical Analysis

The progression of HN target contouring skills among RO RPs was assessed throughout their postgraduate year (PGY) training. Mean contour grades were calculated for each RP and compared throughout PGY progression and residency, with the expectation of improved scores by the end of training. Mean contour grades were also calculated by PGY cohort, with a similar expectation of improvement in the later years of training. Mean contour grades were compared across PGY levels to evaluate correlation. Confidence intervals (CIs) were calculated using the Student $t$ test. Statistical analyses were performed using Prism version 10 (GraphPad Software, Boston, MA).

### Results

Over the course of this study, 218 HN cancer patient targets were contoured by 6 RO RPs and then formally peer reviewed. Among the patient population,

26% were females and 74% were males, with a median age of 67 years. The 5 most common HN tumor sites were oropharynx (35%), cutaneous (18%), oral cavity (12%), salivary glands (8%), and larynx/hypopharynx (6%). Patient characteristics are detailed in **Table 1**.

Of the 218 cases included, 22% (48) were contoured by PGY2 RPs, 45% (98) by PGY3 RPs, 18% (40) by PGY4 RPs, and 15% (32) by PGY5 RPs. An objective improvement in contour grades was observed across advancing training years, with lower scores (trending towards zero) indicating less need for target volume edits or modifications. The mean contour grades for RPs (**Figure 1**) were, for PGY2s, 1.43 (SD = 0.71; CI = 0.2); for PGY3s, 0.99 (SD = 0.81; CI = 0.16); for PGY4s, 0.93 (SD = 0.92; CI = 0.29); and for PGY5s, 0.69 (SD = 0.64; CI = 0.23).

Subsequently, the study assessed the mean contour grade for each RP throughout training (**Figure 2**): resident 1 (1.05, PGY3; 0.5, PGY4; 0.88, PGY5); resident 2 (1.07, PGY3; 1.13, PGY4; 0.75, PGY5); resident 3, (1.38, PGY2; 0.47, PGY3; 0.67, PGY4; 0.55, PGY5); resident 4 (1.30, PGY2; 1.25, PGY3; 1.14, PGY4; 0.70, PGY5); resident 5 (2.00, PGY2; 1.43, PGY3; 0.67, PGY4); and resident 6 (1.06, PGY2; 0.5, PGY3 [currently in training]). Overall, we observed a consistent improvement in contour grades for each RP, with an absolute mean improvement of –0.2 for resident 1, –0.32 for resident 2, –0.82 for resident 3, –0.4 for resident 4, –1.33 for resident 5, and –0.56 for resident 6 across years of training (**Figure 2**).

### Discussion

#### Practice Brings Improvement

This single-institution, prospective study demonstrates that target contour grading is an effective tool for evaluating the progress of RPs' competency in contouring throughout training. A steady reduction in the frequency of R2 contour grades was observed throughout postgraduate training from PGY2 to PGY5, represented by lower mean scores. Meanwhile, an absolute

### Table 1. Patient Characteristics

| CHARACTERISTICS | N (%) OR MEDIAN (RANGE) |
|---|---|
| Gender | |
| Female | 50 (26%) |
| Male | 162 (74) |
| Age | 67 (29-92) |
| Primary site | |
| Oropharynx | 76 (35) |
| Cutaneous | 40 (18) |
| Oral cavity | 25 (12) |
| Salivary glands | 17 (8) |
| Larynx/hypopharynx | 14 (6) |
| Other | 46 (21) |
| Intent | |
| Curative | 209 (96) |
| Palliative | 9 (4) |
| T stage | |
| T0 | 20 (9) |
| T1 | 30 (14) |
| T2 | 33 (15) |
| T3 | 49 (23) |
| T4 | 73 (33) |
| NA | 13 (6) |
| N stage | |
| N0 | 73 (33) |
| N+ | 134 (62) |
| NA | 11 (5) |
| M stage | |
| M0 | 202 (93) |
| M1 | 5 (2) |
| NA | 11 (5) |
| Radiation Therapy (RT) modality | |
| Proton | 122 (56) |
| Intensity Modulated Radiotherapy (IMRT) | 95 (44) |
| 3 Dimensional Conformal Radiation Therapy (3D-CRT) | 1 (0) |
| RT total dose | 60 (28-74.4) |
| RT n of fractions | 30 (12-37) |

improvement in the frequency of individual R0 grades over residency training was observed based on the consensus of expert peer reviewers. This affirms the expectation that as RO RPs progress through training, their contours that would be considered unacceptable for use in treatment decrease, while the number of cases done accurately, requiring no additional modifications, increases over time. Additionally, by PGY4 and PGY5, the frequency of R0 versus R2 grades among residents was consistent with those observed among faculty, supporting the readiness of the RPs for independent HN contouring by graduation.

For some RPs, we observed a nonlinear progression of mean contour grades throughout their training. This trend might be attributed to several factors, including individual learning curve progression; varying levels of experience due to uneven case distribution over the PGY; heterogeneity in case complexity, which was not controlled for according to the RP's PGY; and the well-documented subjectivity associated with volume delineation in HN cancers. Despite this variation, all RPs demonstrated improvement in contouring skills when comparing the beginning of training to the end. This suggests that the institution's peer-review process, and factors such as feedback, documentation, and accountability, have a significant, positive impact on the RP's development of independent contouring skills. Overall, we demonstrate that contour grading allows evaluators and learners to effectively document progression of skills throughout training.

In our institution's timeline of clinical rotations for RO residents, there is a significant increase in the caseload of HN cancer patients between PGY2 and PGY3. In this study, 22% of the cases were contoured by PGY2s and 45% by PGY3s. The reason for this discrepancy is that RPs in the PGY2 HN rotation see both patients with HN cancer and with breast cancer, while those in the PGY3 HN rotation see primarily HN patients. Concurrent to this caseload increase, we observed the largest difference in mean contour grade between PGY2 and PGY3,

**Figure 1.** Contour grading change over the training period. PGY, postgraduate year.
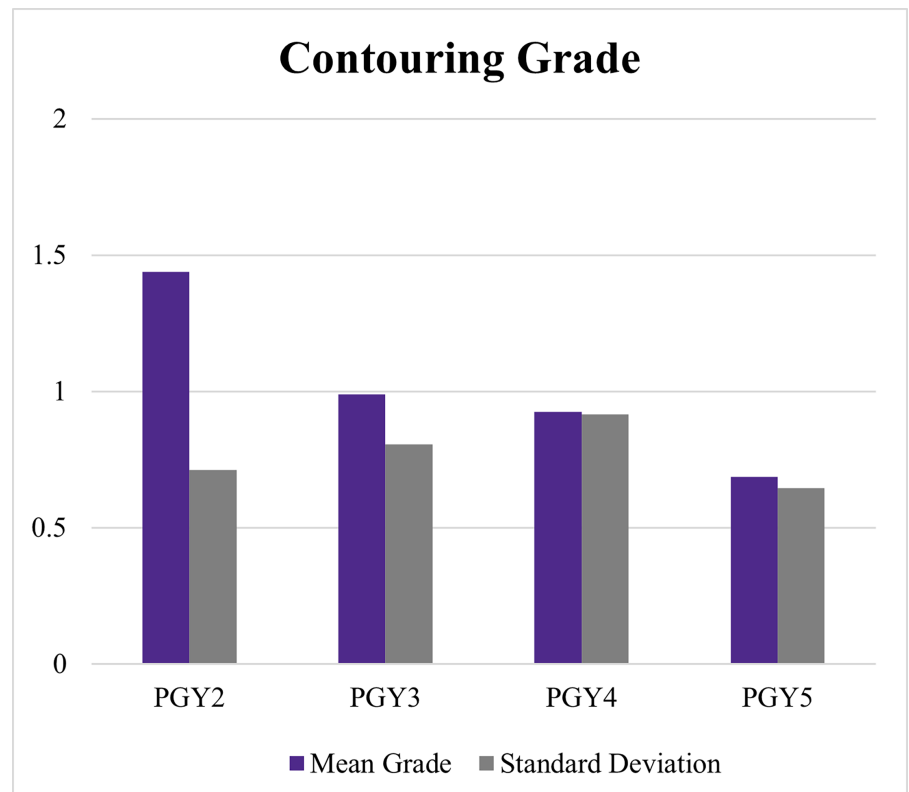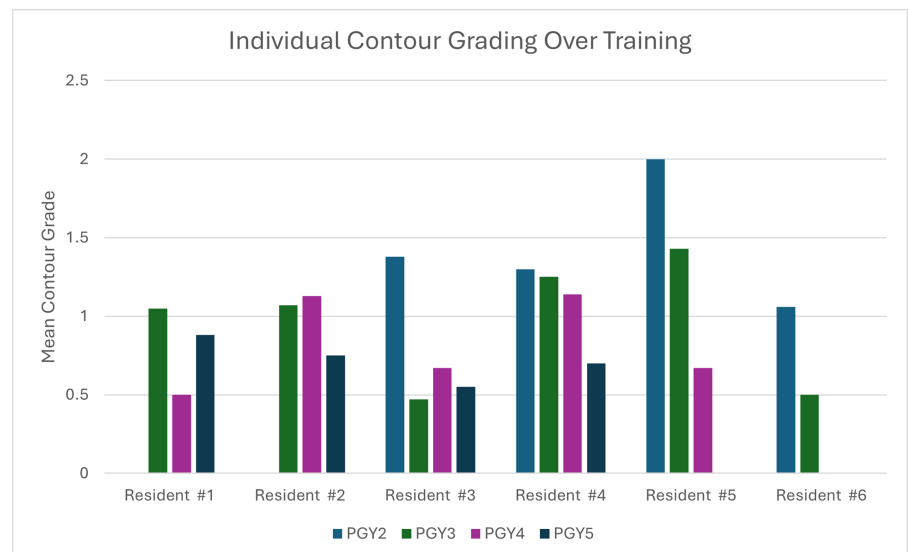


**Figure 2.** Individual contour grading change over the training period. PGY, postgrad uate year.



which suggests that increased contour practice is a significant driver in contouring skill evolution. By PGY3, individuals were more likely to receive a score of R0 or R1 than R2, a trend that is consistent with similar studies revealing a correlation between a physician's contouring skills and level of expertise[7,14]

### Physician Collaboration Improves Treatment Planning

Contour grading, when used in conjunction with peer review, allows for RO RPs to collaborate with more experienced physicians in HN radiation treatment planning (RTP) and evaluation. The process facilitates discussions between APs and RPs that not only address contour improvement skills, but also areas of controversy within RO contouring practice, and different approaches to contouring and treatment based on a patient's anatomy and case specifics. These discussions culminate in consensus among RO APs, which minimizes interobserver variability in practice and decreases the frequency of systemic errors, such as dose delivery or geographic misses, which are known to compromise local control and increase morbidity.[6] This has high educational benefit for RO RPs by presenting more covert viewpoints and emerging considerations, ensuring a well-rounded approach to RTP.

A cross-sectional analysis of recent RO residency graduates in 2016-2017 observed that an increased caseload and independent treatment planning during residency correlated with greater confidence and comfort during independent clinical practice.[18] This is particularly relevant given the findings of a recent needs assessment conducted at 14 ACGME-accredited RO residency programs. In this study, 56% of RPs reported inadequate exposure to RTP, and 54% expressed a lack of confidence in independently evaluating RTP. Additionally, 47% indicated that their education in this area was insufficient, while 97% of all respondents believed that a structured RTP review process could improve RP competency in plan evaluation.[19] The inclusion of contour grading within a similar peer-review framework may serve as a standardized approach to addressing this educational gap, and for better preparing RPs for the transition to independent practice, both in terms of technical skills and mental readiness.

### Opportunities for Objective Performance Measurement

To our knowledge, this is the first prospective study to report the utility of target contour grading as a longitudinal, objective assessment of contouring skill progression in RO RPs. It was designed to address the current lack of objective metrics within the national ACGME RO RP evaluation framework.

ACGME has been transparent with its aspirations to follow the model of graduate education by moving toward a competency-based system of evaluation for residency programs. Its Milestones 1.0, which outlines 6 key competencies and additional, disease-site-specific subcompetencies, has been widely criticized for being difficult to implement consistently, ambiguous differentiation between levels of progression, and prioritization of competencies over key clinical skills. Milestones 2.0, the revised framework released in July 2022,[2] addresses some of these criticisms with the inclusion of an implementation guide and primary goals that focus on clinical skill presentation at each level. While improved, the framework fails to provide objective metrics for key skills such as target volume delineation, making standardization and nationwide comparison of RO residency programs challenging.[1,3] The primary goals remain largely subjective, and the accompanying implementation guides are rarely referenced in RO due to the diversity of the field its cases.

Under the standard process, faculty members evaluate the performance of RPs every 6 months. Without a national, standardized method for evaluating individual cases, this process often results in a generalized, subjective assessment of an RP's abilities rather than a clear, objective measure of skill improvement.

Contour grading offers a solution to this gap by enabling case-by-case scoring that allows institutions to objectively quantify an RP's progression over time. Incorporating objective contour grading into the standardized ACGME

RO RP assessment would help ensure that residents demonstrate measurable proficiency in contouring prior to graduation, thereby preparing them for independent clinical practice.

### Study Limitations

There are some limitations to our study. It was restricted to a single RO department, with a small sample size of 6 RPs, which could potentially limit the applicability of these findings to all RO departments. Additionally, due to the timeframe restriction of reporting, the data do not take into account all of the years of training for each of the 6 included RPs. Longer follow-up would have ensured that more residents who had completed all 4 years of training (PGY2-PGY5) would have been included. Another limitation is the subjectivity of applying a grade to individual cases. Capturing the specific recommendations for contour amendment unique to each case could elucidate systematic errors that could be addressed with curriculum modification.

Incorporating a formal, consensus-driven RTP review process may further strengthen HN peer review and enhance the educational experience for RO RPs. Future research can expand upon these findings to optimize the use of HN peer review as an educational tool, guiding the development of targeted training strategies, educational resources, and objective assessment methods for RO RPs nationwide.

### Conclusion

This study demonstrates that the incorporation of a formal HN contouring peer-review process and RO RP target contour grade assignment into routine clinical practice is feasible and practical. The peer-review process can be used to objectively monitor RO RP contour competency progression and can enhance the existing framework of ACGME milestones.

## References

1) Rajeev-Kumar G, Manjunath R, Hasan Y. Radiation oncology resident evaluations: current methods and resident perceptions. *Adv Radiat Oncol*. 2023;8(5):101230. doi:10.1016/j.adro.2023.101230

2) Thomas HR, Braunstein SE, Golden DW, et al. Milestones 2.0: refining the competency-based assessment of radiation oncology medical residents. *Int J Radiat Oncol Biol Phys*. 2023;115(1):12-18. doi:10.1016/j.ijrobp.2022.07.1838

3) Nissen C, Ying J, Kalantari F, et al. A prospective study measuring resident and faculty contour concordance: a potential tool for quantitative assessment of residents' performance in contouring and target delineation in radiation oncology residency. *J Am Coll Radiol*. 2024;21(3):464-472. doi:10.1016/j.jacr.2023.08.049

4) American Society for Radiation Oncology (ASTRO). RO-ILS case study 10: contour delineation: a crucial treatment planning process. Accessed January 3, 2025. https://www.astro.org/practice-support/quality-and-safety/ro-ils/case-studies/case-study-10

5) Abugideiri M, Schreibmann E, Switchenko J, et al. Prospective international pilot study evaluating the efficacy of a self-guided contouring teaching module with integrated feedback for transitioning from 2D to 3D treatment planning. *J Glob Oncol*. 2019;5(5):1-16. doi:10.1200/JGO.18.00224

6) Segedin B, Petric P. Uncertainties in target volume delineation in radiotherapy - are they relevant and what can we do about them? *Radiol Oncol*. 2016;50(3):254-262. doi:10.1515/raon-2016-0023

7) Gogineni E, Schaefer D, Ewing A, et al. Systematic implementation of effective quality assurance processes for the assessment of radiation target volumes in head and neck cancer. *Pract Radiat Oncol*. 2024;14(3):e205-e213. doi:10.1016/j.prro.2023.12.012

8) Bekelman JE, Wolden S, Lee N. Head-and-neck target delineation among radiation oncology residents after a teaching intervention: a prospective, blinded pilot study. *Int J Radiat Oncol Biol Phys*. 2009;73(2):416-423. doi:10.1016/j.ijrobp.2008.04.028

9) Martin-Garcia E, Celada-Álvarez F, Pérez-Calatayud MJ, et al. 100% peer review in radiation oncology: is it feasible? *Clin Transl Oncol*. 2020;22(12):2341-2349. doi:10.1007/s12094-020-02394-8

10) Gwynne S, Spezi E, Sebag-Montefiore D, et al. Improving radiotherapy quality assurance in clinical trials: assessment of target volume delineation of the pre-accrual benchmark case. *Br J Radiol*. 2013;86(1024):20120398. doi:10.1259/bjr.20120398

11) Nielsen CP, Lorenzen EL, Jensen K, et al. Interobserver variation in organs at risk contouring in head and neck cancer according to the DAHANCA guidelines. *Radiother Oncol*. 2024;197:110337. doi:10.1016/j.radonc.2024.110337

12) Nuyts S. Defining the target for radiotherapy of head and neck cancer. *Cancer Imaging*. 2007;7 Spec No A(Special issue A):S50-S55. doi:10.1102/1470-7330.2007.9009

13) Marks LB, Adams RD, Pawlicki T, et al. Enhancing the role of case-oriented peer review to improve quality and safety in radiation oncology: executive summary. *Pract Radiat Oncol*. 2013;3(3):149-156. doi:10.1016/j.prro.2012.11.010

14) Ballo MT, Chronowski GM, Schlembach PJ, et al. Prospective peer review quality assurance for outpatient radiation therapy. *Pract Radiat Oncol*. 2014;4(5):279-284. doi:10.1016/j.prro.2013.11.004

15) Hoopes DJ, Johnstone PA, Chapin PS, et al. Practice patterns for peer review in radiation oncology. *Pract Radiat Oncol*. 2015;5(1):32-38. doi:10.1016/j.prro.2014.04.004

16) Duggar WN, Bhandari R, Yang CC, Vijayakumar S. Group consensus peer review in radiation oncology: commitment to quality. *Radiat Oncol*. 2018;13(1):55. doi:10.1186/s13014-018-1006-1

17) Cooper BT, Goenka A, Sine K, et al. Development of a comprehensive, contour-based, peer review workflow at a community proton center. *Int J Part Ther*. 2020;7(1):34-40. doi:10.14338/IJPT-19-00059.1

18) Brower JV, Chen S, Ritter A, et al. Comfort level of US radiation oncology graduates: assessment of transition to independent clinical practice. *J Cancer Educ*. 2021;36(2):278-283. doi:10.1007/s13187-019-01625-z

19) Boyd GH, Vanbenthuysen L, Jimenez RB. Radiation treatment plan evaluation education in residency: a needs assessment. *Int J Radiat Oncol Biol Phys*. 2023;117(2):e504-e505. doi:10.1016/j.ijrobp.2023.06.1753