



Published in final edited form as:

Int J Radiat Oncol Biol Phys. 2017 July 01; 98(3): 532–540. doi:10.1016/j.ijrobp.2016.11.019.

Prospective Qualitative and Quantitative Analysis of Real-time Peer Review Quality Assurance Rounds Incorporating Direct Physical Examination for Head and Neck Cancer Radiation Therapy

Carlos E. Cardenas^{*}, Abdallah S.R. Mohamed^{†,‡}, Randa Tao[†], Andrew J.R. Wong[†], Musaddiq J. Awan[†], Shirly Kuruvila[†], Michalis Aristophanous^{*}, G. Brandon Gunn[†], Jack Phan[†], Beth M. Beadle[†], Steven J. Frank[†], Adam S. Garden[†], William H. Morrison[†], Clifton David Fuller[†], and David I. Rosenthal[†]

^{*}Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas

[†]Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas

[‡]Department of Clinical Oncology and Nuclear Medicine, Faculty of Medicine, University of Alexandria, Alexandria, Egypt

Abstract

Purpose and Objectives—Our department has a long-established comprehensive quality assurance (QA) planning clinic for patients undergoing radiation therapy (RT) for head and neck cancer. Our aim is to assess the impact of a real-time peer review QA process on the quantitative and qualitative radiation therapy plan changes in the era of intensity modulated RT (IMRT).

Methods and Materials—Prospective data for 85 patients undergoing head and neck IMRT who presented at a biweekly QA clinic after simulation and contouring were collected. A standard data collection form was used to document alterations made during this process. The original pre-QA clinical target volumes (CTVs) approved by the treating-attending physicians were saved before QA and compared with post-QA consensus CTVs. Qualitative assessment was done according to predefined criteria. Dice similarity coefficients (DSC) and other volume overlap metrics were calculated for each CTV level and were used for quantitative comparison. Changes are categorized as major, minor, and trivial according to the degree of overlap. Patterns of failure were analyzed and correlated to plan changes.

Reprint requests to: David I. Rosenthal, MD, Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd., Unit 97, Houston, TX 77030. Tel: (713) 563-2353; dirosenthal@mdanderson.org or Clifton D. Fuller, MD, PhD, Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd., Unit 97, Houston, TX 77030. Tel: (713) 563-2334; cdfuller@mdanderson.org.

Presented in part as an oral presentation at the 54th Annual Meeting of the American Society for Radiation Oncology, October 31, 2012, Boston, MA.

C. E. Cardenas, A. S. R. Mohamed, and R. Tao contributed equally to this study.

Conflict of interest: none.

Supplementary material for this article can be found at www.redjournal.org.

Results—All 85 patients were examined by at least 1 head and neck subspecialist radiation oncologist who was not the treating-attending physician; 80 (94%) were examined by 3 faculty members. New clinical findings on physical examination were found in 12 patients (14%) leading to major plan changes. Quantitative DSC analysis revealed significantly better agreement in CTV1 (0.94 ± 0.10) contours than in CTV2 (0.82 ± 0.25) and CTV3 (0.86 ± 0.2) contours ($P=.0002$ and $P=.03$, respectively; matched-pair Wil-coxon test). The experience of the treating-attending radiation oncologist significantly affected DSC values when all CTV levels were considered ($P=.012$; matched-pair Wil-coxon test). After a median follow-up time of 38 months, only 10 patients (12%) had local recurrence, regional recurrence, or both, mostly in central high-dose areas.

Conclusions—Comprehensive peer review planning clinic is an essential component of IMRT QA that led to major changes in one third of the study population. This process ensured safety related to target definition and led to favorable disease control profiles, with no identifiable recurrences attributable to geometric misses or delineation errors.

Introduction

Extensive guidelines are available on the approaches to the physical and technical aspects of quality assurance (QA) to ensure accurate and safe delivery of radiation therapy (RT), but there are fewer standardized approaches to clinical QA (1-3). Traditionally, clinical QA is accomplished during chart rounds, in which the RT plan and perhaps diagnostic imaging are reviewed. With the evolution of RT toward more conformal treatment of physician-defined target volumes and organs at risk (OARs), clinical QA and physician experience (4, 5) have become recognized as increasingly important to patient care along with the critical QA processes established for treatment machines in radiation physics (6).

At the University XXX's radiation oncology department, we have had a head and neck planning and development clinic (HNPDC) since its establishment by Dr XXX in the 1960s. The main goals are comprehensive peer review of a patient's RT plan before the start of RT and trainee teaching (7). In addition to the comprehensive review of a patient's history, pathologic characteristics, diagnostic imaging, and discussion of the planned treatment, all patients undergo physical examination (PE), including videocamera naso-pharyngolaryngoscopy and bimanual palpation performed by the head-and-neck radiation oncology subspecialists in our group. The proposed computed tomographic (CT) image is reviewed slice by slice for gross tumor volume (GTV), clinical target volume (CTV), OAR segmentation, and dose-volume specifications.

We previously examined the value of this peer review process to characterize the types of changes proposed during the QA process. Changes in the radiation plans were noted in the majority of patients included in that study. Of those, 11% were considered major changes, defined as potentially affecting patient outcomes (7). The majority of RT plans in that study were 3-dimensional (3D) conformal, and we demonstrated that PE is critical in the process of target volume delineation in those cases. Currently, in the era of intensity modulated radiation therapy (IMRT) and proton therapy (IMPT) with their high level of treatment precision, target volume definition is increasingly critical, particularly in head and neck cancer (8, 9). As a continuation of efforts to characterize the impact of a real-time peer

review QA process, we performed an update of that previous study to assess the changes made during peer review in the setting of exclusive IMRT planning. A quantitative analysis of the volumetric changes made as a result of HNPDC and patterns of failure analysis are included in this study.

Methods and Materials

Patients and data collection

Data were collected from January to May 2012 from 85 consecutive patients who presented at the HNPDC, using a standard data collection form by 2 individuals in our head and neck group. The qualitative patient information recorded for each case included patient and tumor characteristics, treating-attending physician, attending physicians performing peer review, and new findings by the group. Group recommendations such as suggesting additional tests, changes in chemotherapy plan, and changes in the proposed radiation volume, dose, or fractionation were also recorded.

The HNPDC peer review process, which follows our institution's clearly defined treatment policies (10), has been summarized in detail by XXX et al (7). The process can be described as follows: patients undergo a CT-based simulation under the care of the treating-attending physician. Before the HNPDC, the treating-attending physician and the resident trainee contour the CTVs and OARs using Pinnacle (Philips Healthcare, Andover, MA) or Eclipse (Varian Medical Systems, Palo Alto, CA) treatment planning software. The high-risk CTV (CTV1) is defined as the GTV plus margin in definitive cases or the tumor bed plus margin in postoperative cases. The intermediate-risk CTV (CTV2) is then defined as the adjacent region to the high-risk CTV, and a third CTV3 includes an elective region such as elective nodal basins or contralateral neck. After the delineation of initial target volumes by the attending radiation oncologist, plans are subsequently presented at the HNPDC, usually 1 to 2 days after the date of simulation. Each patient scheduled to undergo curative-intent RT is present during the HNPDC. As each patient's case is presented, the treating radiation oncologist summarizes details of the patient's history and tumor characteristics along with a review of the patient's diagnostic imaging. The peer group then performs a PE, which includes inspection, palpation (including bimanual examination to define the submucosal borders of a tongue cancer) and a video-camera nasopharyngolaryngoscopy viewed by the group on a video monitor. Finally, the nontreating physicians conduct a peer review of the contoured volumes of the target and the OARs. Suggested changes are discussed until a consensus is reached. Real-time changes are made to target the volume contours during the review process. Once all required changes are made and a consensus is reached, the plan is submitted for dosimetric planning. Physics QA is performed before the start of treatment.

Eight subspecialist board-certified/eligible head and neck radiation oncologists participated in the study. Two were junior faculty physicians who had less than 2 years of posttraining experience and were defined as “less experienced” in our analysis. The remaining 6 were defined as “experienced.”

Qualitative assessment

As in our institution's previous study (7), changes in treatment plan were qualitatively classified as major if they were believed to clinically affect the likelihood of cure, adverse events, or locoregional control. Changes were considered minor when the recommendations made were more elective or stylistic; they included modifications of field delineation for additional margins to accommodate penumbra or potential motion/position change, and changing the fractionation schedule.

Quantitative metrics

The treating physician's original (pre-HNPDC) target volumes (CTV_{pre}) were saved in Pinnacle to facilitate a quantitative analysis of pre-HNPDC and post-HNPDC target volume contours. The final (post-HNPDC) contours (CTV_{QA}) submitted after peer review were then compared with the initial contours by use of volume-based measures. Target volumes (cm^3) were computed with the use of Pinnacle for pre-HNPDC and post-HNPDC contours, and their volumetric difference (VD) was calculated with Eq. 1:

$$VD = \frac{V_{pre} - V_{QA}}{V_{QA}} \quad (1)$$

To characterize the spatial overlap between the CTV_{pre} and CTV_{QA} volumes, the Dice similarity coefficient (DSC) was calculated with the use of Eq. 2:

$$DSC = \frac{2 |V_{pre} \cap V_{QA}|}{|V_{pre}| + |V_{QA}|} \quad (2)$$

In addition, the false negative Dice (FND) and the false positive Dice (FPD) were calculated to assess potential near misses and overtreatment, respectively. FND (Eq. 3) measures the volume added after peer review that was originally not included by the treating-attending physician, and FPD (Eq. 4) measures the volume removed after peer review that the treating-attending originally included. Figure E1 (available online at www.redjournal.org) shows a diagram illustrating these metrics, and Figure 1 demonstrates an example of 2 patients with post-QA CTV changes and their effect on the FND and FPD overlap metrics.

$$FND = \frac{2 |V_{pre} \cap V_{QA}|}{|V_{pre}| + |V_{QA}|} \quad (3)$$

$$\text{FPD} = \frac{2 |V_{\text{pre}} \cap \overline{V_{\text{QA}}}|}{|V_{\text{pre}}| + |V_{\text{QA}}|} \quad (4)$$

Patterns of failure analysis

Patients completing treatment received initial posttreatment evaluations after 8 to 12 weeks. Subsequently, patients were followed up every 2 to 3 months for the first year, every 3 to 4 months for the second year, and at least twice a year up to 5 years. Ten patients with documented posttreatment recurrence were analyzed for the relation of volume recurrence relative to both pre-HNPDC and post-HNPDC volumes. The CT/PET-CT images documenting recurrence were manually delineated by a single observer (XXX) and reviewed by 2 head and neck radiation oncologists (XXX, XXX). The recurrence images were subsequently registered to the planning CT with deformable image registration software Velocity AI 3.0.1 (Velocity Medical, Atlanta, GA). The recurrence contours were reviewed relative to the spatial location of the original and post-QA contours and to the dose grid according to our institutional patterns of failure classification scheme (11).

Statistical analysis

The Pearson χ^2 test was used for comparing qualitative assessments according to different HNPDC associated covariates. Statistical comparison of the volumetric changes in CTV levels was performed with the paired Wilcoxon rank sum test. In addition, quantitative metrics were classified by threshold values, as shown in Table E1 (available online at www.redjournal.org). Although the literature recommends a DSC >0.70 for good overlap (12, 13), we took a more conservative approach in our analysis by slightly raising these threshold values. The overall DSC change was determined for each patient as follows: if any change in CTV level was classified as major, then the overall DSC change was labeled major; if no changes in CTV levels were classified as major but at least 1 change was classified as minor, then the overall DSC was labeled as minor; only when all changes in CTV levels were deemed trivial was the overall DSC labeled trivial. Data analysis was performed with the JMP version 12.1 statistical software (SAS Institute, Cary, NC).

Results

Qualitative plan changes and physical examination

The patient and tumor characteristics are summarized in Table 1. Of the 85 patients, 37 received definitive IMRT, and 48 received postoperative IMRT. All patients were examined by at least 1 nontreating attending radiation oncologist specializing in head and neck cancers, and 80 were examined by 3 physicians (94%). The median number of attending physicians present at the HNPDC (including the treating-attending) was 4. The CT images of the head and neck were reviewed for all patients; additional review of images included positron emission tomography (PET)/CT in 17 patients (20%), magnetic resonance imaging (MRI) in 3 patients (3.5%), and both PET/CT and MRI in 3 patients (3.5%). Group videcamera nasopharyngolaryngoscopy was performed on 37 (44%) patients, including

70% of patients with larynx primaries, 100% of nasopharynx primaries, 5% of oral cavity patients, 92% of oropharynx patients, and 67% of patients with unknown primaries. Endoscopy was not performed on patients with skin, salivary gland, and thyroid primaries.

New findings as a result of group PE during HNPDC were documented for 12 patients (14%). These findings included evidence of tumor progression (n=3), extension to surrounding anatomic structures such as soft palate, aryepiglottic fold, or piriform sinus (n=3), further tumor shrinkage after chemotherapy (n=3), impaired vocal cord mobility (n=2), and new neck adenopathy (n=1). Of these, 7 of 12 findings (58%) were identified by group bimanual palpation and the remainder by endoscopic examination. In all cases, no changes were suggested to reject the use of RT or to change the treatment intent from curative to palliative.

The IMRT plans featured qualitative changes in 65 cases (76%); 30 patients (35%) were classified as having major changes, and 35 patients (41%) were classified as having minor changes. There was no relationship between endoscopic examination and the frequency or type of qualitative changes ($P=.74$); furthermore, the site of the primary tumor was not associated with differences in the frequency or type of changes made ($P=.22$). No difference was seen in the frequency or type of qualitative changes between patients treated with definitive IMRT and those treated postoperatively ($P=.77$). Tumor stage did not have a relationship with frequency or type of qualitative change ($P=.78$). There was no relationship between the number of attending physicians present and the frequency or type of qualitative changes ($P=.12$). Statistically significant higher rates of both major and minor changes were noted in the junior faculty plans compared with plans made by more experienced faculty ($P<.0001$), with a rate of 69% major changes in junior faculty plans compared with 20% for more experienced faculty plans. However, there was not a significant difference between the individual members of each group (junior vs more experienced faculty) and the frequency or type of qualitative change (experienced, $P=.30$; junior, $P=.12$).

A change in the total planned IMRT dose was suggested for 9 patients (11%); 8 of them were recommended to receive an increase in the prescribed dose to the primary and nodal regions, and 1 patient received a reduced dose. The addition or subtraction of a dose level was suggested for 12 patients (14%). The addition of electron fields was suggested for 3 patients. One patient's plan changed from a whole field IMRT technique to a matched IMRT field with an anteroposterior supraclavicular field, whereas the opposite was done for another patient. Last, the addition of contralateral lymph node coverage was suggested for a single patient.

Quantitative analysis of clinical target volumes

Differences in percentage for each type of qualitative and quantitative changes according to physician experience are illustrated in Figure 2. One patient was excluded from this analysis because of the addition of 2 dose levels to the post-HNPDC treatment plan. Quantitative results and analysis of metrics according to CTV level are summarized in Table 2. When the overall DSC classification was considered for all CTV levels, 23 patients (27%) had major modifications, and 36 (43%) had minor changes. The same metric showed that junior physicians had a lower rate of trivial changes (12%) than did more experienced physicians

(37%, $P=.02$) and had a higher rate of major changes (44% junior vs 20% experienced, $P=.03$). Major change rates were also found to be statistically significant when overall FPD was considered (48% junior vs 24% experienced, $P=.03$). Figure 3 depicts the box plots of each individual overlap metric by experience and dose levels.

When considering DSC values for CTV1 were considered, no relationship was observed between DSC value and endoscopic examination ($P=.82$). Other covariates such as number of attending physicians ($P=.30$), tumor site ($P=.22$), and tumor stage ($P=.79$) were not found to have a significant impact on the frequency of overall DSC changes. A higher rate of major changes was noted in patients receiving definitive RT (36%) than in patients treated postoperatively (21%), but this difference was not statistically significant ($P=.12$).

The resulting DSC values for CTV1 for all physicians were significantly higher than those for CTV2 and CTV3 (0.94 ± 0.12 vs 0.84 ± 0.23 , and 0.87 ± 0.20 ; $P=.0005$ and $P=.03$, respectively) but no difference was observed in DSC values between CTV2 and CTV3 contours ($P=.23$). When physician experience was considered, the DSC values for CTV1 were higher than for CTV2 in both groups (experienced, 0.96 ± 0.05 vs 0.88 ± 0.19 , $P=.02$; junior, 0.92 ± 0.17 vs 0.76 ± 0.26 , $P=.005$). The DSC values from CTV1 and CTV3 showed higher agreement for the experienced physicians (0.96 ± 0.05 vs 0.90 ± 0.19 , $P=.53$), whereas for the junior physicians, the CTV3 values were significantly lower (0.92 ± 0.17 vs 0.80 ± 0.25 , $P=.009$). The CTV2 and CTV3 DSC values showed no difference for the junior physicians (0.75 ± 0.26 vs 0.80 ± 0.25 , $P=.81$), but for the experienced group, this difference was slightly increased (0.88 ± 0.18 vs 0.90 ± 0.19 , $P=.11$), with the CTV3 DSC values being higher. Overall, physician experience was correlated with higher DSC values when all CTV levels were considered (0.91 ± 0.16 vs 0.83 ± 0.23 , $P=.005$). Table 3 shows a summary of quantitative changes classification. The rate of FND major changes in CTV1 was significantly higher than FPD major changes (16% vs 2%, $P=.003$), suggesting that changes in CTV1 contours were more likely to include additional coverage, possibly as a result of tumor progression or further extension findings noted during the real-time peer review QA clinic.

Patterns of failure

The median follow-up time was 38 months (range, 3-49 months). Two patients did not receive the planned RT after HNPDC, leaving 83 patients for patterns of failure analysis; 1 patient decided to postpone treatment because she was pregnant, and another patient's disease metastasized to the lungs and the patient refused any further treatment. Nine patients had documented local recurrence, and 1 patient had regional recurrence. Seven patients had a central high-dose failure (ie, the entire failure volume is encapsulated by the 95% isodose line and entirely within CTV1). Three patients had noncentral high-dose recurrences. Of those, 1 patient had a central intermediate dose failure (ie, entirely within CTV2). This recurrence was in the right buccal mucosa 2 years after treatment for T4N0 squamous cell carcinoma of the right floor of the mouth with surgical resection followed by flap reconstruction and adjuvant IMRT. The second patient with noncentral high-dose recurrence had a peripheral high-dose failure (ie, the failure centroid was marginal to CTV1, and the failure volume had less than the 95% dose). This patient was treated for left sinonasal

recurrence (prior surgery but not prior radiation), and had progressive intracranial extension immediately after IMRT. Last, the third patient with noncentral recurrence experienced extraneous failure in the contralateral right level II-IV neck 3 months after ipsilateral postoperative IMRT, after salvage surgical resection of right parotid recurrence in a patient with a history of T2 N0 squamous cell carcinoma of the left alveolar ridge.

The target volumes before and after HNPDC did not affect the definition of patterns of failure for both central and noncentral high-dose failures (ie, the recurrence volume carried a similar spatial relative location to both the pre-HNPDC and post-HNPDC target volume contours).

Discussion

Our institution's HNPDC peer review QA continues as a comprehensive and rigorous process, whose main goals are preventing tumor misses, preventing unnecessary normal structure treatment/dosing, and reducing operator error. The patient examination and case reviews in this study revealed that 35% of head and neck treatment plans required major qualitative changes, and 30% of patient CTV contours required major changes during the HNPDC. It is our belief that without these changes the rates of cure could have been compromised and that the results presented in this study validate the need of a comprehensive weekly QA program.

We have previously reported that in about 10% of cases, major qualitative changes were recommended under the same guidelines used in this study (7). However, in the current study, the rate of qualitative major changes was 35%. The relatively higher rate of major qualitative changes in this study in comparison with the previous study reflects the requirement of more rigorous QA of IMRT plans compared with 3D conformal plans, which were the majority of plans in the previous study (7). Additionally, the contribution of more junior faculty in this study, with their higher rates of major plan modification, can also explain the same findings.

Quantitatively, more experienced faculty had fewer post-QA CTV changes as measured by all metrics than did less experienced faculty, concurring with the qualitative assessment and highlighting the crucial role of this peer review QA process, particularly for less experienced physicians. Additionally, the overall CTV volumes of the postoperative plans were shown to have fewer volumetric changes compared with the CTV volumes of the definitive plans. This is likely due to the already recognized larger volumes required for postoperative RT. This significantly higher magnitude of changes in definitive cases may also be due to the effect of the group clinical examination and review of the patient's GTV compared with postoperative cases with no gross tumor left. However, the changes measured in high-dose CTVs for all cases were significantly lower than for intermediate and low-dose CTVs, denoting more agreement in the definition of high-dose target volumes compared with elective or lower-dose target volumes. Differences in the rates of major changes in FND and FPD for CTV1 contours suggested that the majority of changes in this CTV level were possibly due to findings of tumor progression during PE.

The importance of physician-led peer review chart rounds and the quality of RT has been the focus of several recent publications (4, 14, 15). A recent survey revealed that the implementation of clinical QA programs varies greatly across American academic institutions and that the average time spent on each case during chart rounds was 2.7 minutes (14). Overall, the majority of institutions surveyed reported that treatment changes were rare and occurred for <10% of cases presented at chart rounds. Only 11% of institutions reported that major changes were made to more than 10% of cases, and 39% of institutions reported making minor changes. The evaluation of other clinical QA programs has illustrated their value in RT departments. Ballo et al (15) showed that since the inception of their weekly QA program the number of changes recommended for head-and-neck RT plans significantly decreased ($P=.04$) from 44.8% to 26.1% from 2007 to 2010. Additionally, gastrointestinal, gynecologic, and breast cancer decreased significantly over the 4-year period in this same study. These findings suggest that the process of group QA can at least lead to greater uniformity in therapeutic approaches.

Recent publications have addressed the relationship between a center and a radiation oncologist's patient throughput and quality of RT (4, 5, 16). Peters et al (4) suggested that to achieve quality RT in a clinical trial, participation should be limited to clinics that can provide a large number of patients, which is closely related to radiation oncologist experience and subspecialization. This was supported by Wuthrick et al (5), who reported that overall survival, progression-free survival, and locoregional failure were significantly worse when patients were treated at historically low-accruing versus high-accruing cancer centers. Boero et al (16) conducted a retrospective study on the relationship between yearly patient volume per radiation oncologist (using Current Procedural Terminology codes) and patient outcomes. In that study, the authors reported that there was no significant relationship between provider volume and patient survival or any toxicity endpoint among patients treated with conventional radiation; by contrast, patients receiving IMRT by the higher-volume radiation oncologists had improved survival compared with those treated by low-volume providers.

Even though the current study does not address patient survival outcomes, our patterns of locoregional failure analysis showed that despite the rigor of our department's clinical QA program, 12% of patients still experienced local and regional recurrences. These are largely attributed to radioresistance or aberrant patterns of spread rather than to geographic misses, inasmuch as central high-dose failure represented 7 of 10 patients with locoregional failure. These results demonstrate that even though the patients included in the current study had a wide variety of disease primary site, with the majority presenting with locally advanced (74%) or recurrent (16.5%) high-risk disease, only a very few patients ($n=3$) experienced noncentral high-dose failure after IMRT. The analysis of those 3 cases showed that the first patient with intermediate dose failure likely had new primary disease in the ipsilateral buccal mucosa after 2 years of local control after treatment of primary disease in the floor of the mouth. The second patient had the only peripheral high-dose failure caused by progressive intracranial disease extension in a patient treated for recurrent disease after salvage surgery. The third patient had an aberrant pattern of spread in the contralateral neck. This patient initially underwent resection of early T2 alveolar cancer, with no indication for postoperative RT. That patient experienced a recurrence in the ipsilateral parotid and

underwent salvage surgery and ipsilateral postoperative RT but later experienced recurrence in the contralateral neck. All patients with noncentral failure thus had recurrence outside of the standard anatomic patterns of spread. These represented anomalous disease progression that was not attributable to errors in target delineation or to variations in delivery. This supports the utility of these QA efforts in precluding preventable operator-dependent delineation oversights, inasmuch as no geometric misses were observed with a median of >3 years of follow-up.

Our data suggest that even within a large group of subspecialized head and neck radiation oncologists, PE and peer-reviewed QA clinic play a major role in preRT changes, both qualitatively and quantitatively. In addition, the experience of treating-attending physicians showed significant differences in frequency and type of the changes evaluated, which might suggest that chart rounds and clinical QA have pedagogic value for junior radiation oncologists.

Inasmuch as our clinical QA process relies on the availability of a large group of experienced head and neck radiation oncologists, there might be some difficulties translating this approach to smaller clinics. The clinical QA process could be carried out by telemedicine or could be implemented by consulting the referring head and neck oncologists, surgeons, specialists, or a combination of them. The use of telemedicine would lack the critical step of patient examination, namely direct palpation, which would fail to provide a complete evaluation for addressing 3D tumor and target localization. Consequently, it is our belief that any technical efforts (eg, high-quality endoscopic video transfer) that afford improved peer review will enhance the performance of remote assessment by a colleague; however, in the end, the time-tested standard of direct PE remains the benchmark for high-quality head and neck cancer evaluation. Our recommendation, then, is that in the absence of a qualified head and neck radiation oncologist, if at all feasible, the referring surgeon should formally affirm tumor localization, documenting carefully by notation in the electronic medical record, video-endoscopic recordings, or both.

Our study has a few limitations. Because this study was observational, data collection and treatment planning changes were performed in real time during chart rounds, and typical clinical practices remained unchanged. Although the junior physicians could have been less likely to criticize the experienced members' treatment plans and contours, the more experienced physicians constituted a majority (6 of 8 physicians) of the members in the clinical QA process. For those experienced members, approximately 24% of plans required major quantitative changes, showing that all members received a rigorous evaluation of treatment plans. Also, the classification of changes as major and minor were selected empirically according to what the authors deemed appropriate before the start of this study. In addition, it is difficult to assess the ultimate clinical impact of the changes made during the HNPDC because there was no way other to randomize patients to prove that patients benefited from this rigorous peer review process. We believe this process is essential from a safety and practical standpoint, so we believe it would be unethical to randomize patients to an "outside chart review" group versus a group receiving our comprehensive QA, including

PE by HN subspecialists, especially in the light of emerging data about the importance of operator experience.

In conclusion, practicing clinical QA chart rounds offers many benefits to radiation oncology departments. The current study demonstrated that the level of subspecialty experience played a major role in the frequency and type of changes made in patients' IMRT plans. Additionally, thorough PE of patients and review of contours by subspecialty peers are essential components of QA of plans, which led to major changes for approximately one third of the study population. Our belief remains that the HNPDC QA process offers greater value than sole chart and patient imaging review for preventing errors, reducing adverse events, and improving safety related to target localization for patients undergoing IMRT for head and neck cancer, ultimately leading to favorable disease control and very few noncentral high-dose failures. Without adequate peer review of patient treatments, including comprehensive PE, there is an increased risk that target delineation errors could go unnoticed and directly affect the patient outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported in part by the National Institutes of Health (NIH)/National Cancer Institute (NCI) Cancer Center Support (Core) Grant CA016672 to The University of Texas MD Anderson Cancer Center.

References

1. Williamson JF, Dunscombe PB, Sharpe MB, et al. Quality assurance needs for modern image-based radiotherapy: Recommendations from 2007 interorganizational symposium on "quality assurance of radiation therapy: Challenges of advanced technology". *Int J Radiat Oncol Biol Phys.* 2008; 71:2–12. [PubMed: 18406881]
2. Klein EE, Hanley J, Bayouth J, et al. Task group 142 report: Quality assurance of medical accelerators. *Med Phys.* 2009; 36:4197–4212. [PubMed: 19810494]
3. Kutcher GJ, Coia L, Gillin M, et al. Comprehensive QA for radiation oncology: Report of AAPM radiation therapy committee task group 40. *Med Phys.* 1994; 21:581–618. [PubMed: 8058027]
4. 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. [PubMed: 20479390]
5. 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. [PubMed: 25488965]
6. 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; 3:149–156. [PubMed: 24175002]
7. Rosenthal DI, Asper JA, Barker JL, Garden AS, Chao KSC, Morrison WH, et al. Importance of patient examination to clinical quality assurance in head and neck radiation oncology. *Head Neck.* 2006 Nov; 28(11):967–73. [PubMed: 16823872]
8. Eisbruch A, Marsh LH, Dawson LA, et al. Recurrences near base of skull after IMRT for head-and-neck cancer: Implications for target delineation in high neck and for parotid gland sparing. *Int J Radiat Oncol Biol Phys.* 2004; 59:28–42. [PubMed: 15093896]

9. Gupta T, Jain S, Agarwal JP, et al. Prospective assessment of patterns of failure after high-precision definitive (chemo)radiation in head-and-neck squamous cell carcinoma. *Int J Radiat Oncol Biol Phys.* 2011; 80:522–531. [PubMed: 20646862]
10. Ang, KK., Garden, AS. 4th Revise. Philadelphia, United States: Lippincott Williams and Wilkins; 2011. *Radiotherapy for Head and Neck Cancers: Indications and Techniques.*
11. Mohamed ASR, Rosenthal DI, Awan MJ, Garden AS, Kocak-Uzel E, Belal AM, et al. Methodology for analysis and reporting patterns of failure in the Era of IMRT: head and neck cancer applications. *Radiat Oncol Radiation Oncology.* 2016; 11(1):95. [PubMed: 27460585]
12. Zijdenbos AP, Member S, Dawant BM, et al. Morphometric analysis of white matter lesions in MR images: Method and validation. *IEEE Trans Med Imaging.* 1994; 13:16–24.
13. Zou KH, Warfield SK, Bharatha A, et al. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad Radiol.* 2004; 11:178–189. [PubMed: 14974593]
14. Lawrence YR, Whiton MA, Symon Z, et al. Quality assurance peer review chart rounds in 2011: A survey of academic institutions in the United States. *Int J Radiat Oncol Biol Phys.* 2012; 84:590–595. [PubMed: 22445006]
15. Ballo MT, Chronowski GM, Schlembach PJ, Bloom ES, Arzu IY, Kuban DA. Prospective peer review quality assurance for outpatient radiation therapy. *Pract Radiat Oncol American Society for Radiation Oncology.* 2014; 4(5):279–84.
16. 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; 34:685–690.

Summary

Clinical quality assurance (QA) chart rounds vary among institutions. Our head-and-neck group incorporates physical examination during radiation therapy treatment planning peer review. We conducted an observational study to quantify the effect of physical examination and target volume delineation changes during clinical QA. The findings during physical examination lead to changes in approximately one third of cases, and the experience of the treating-attending; physician has been found to significantly affect the frequency and type of treatment plan and target volume changes.

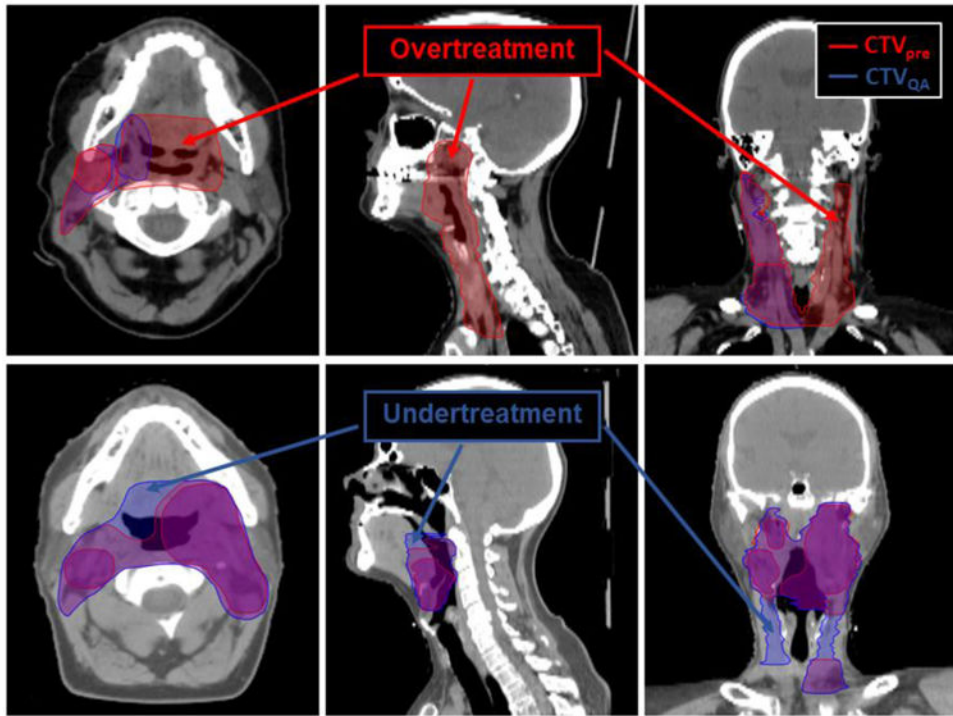


Fig. 1.

Patient images showing differences in clinical target volume (CTV) changes and their effect on false negative Dice (FND) and false positive Dice (FPD) metrics. Top row, simulation CT images for a patient with T1N1 cancer of the right tonsil. Differences in contours before (red) and after (blue) quality assurance (QA) are visualized. For this patient, high FND and Dice similarity coefficient (overtreatment) values were recorded for CTV1 and CTV3, as can be visually assessed on the axial and coronal images. Bottom row, a patient with T4N2c of the left tonsil who was treated with definitive radiation therapy. CTV volumes after QA (blue) show regions that were added and could be considered potential near misses.

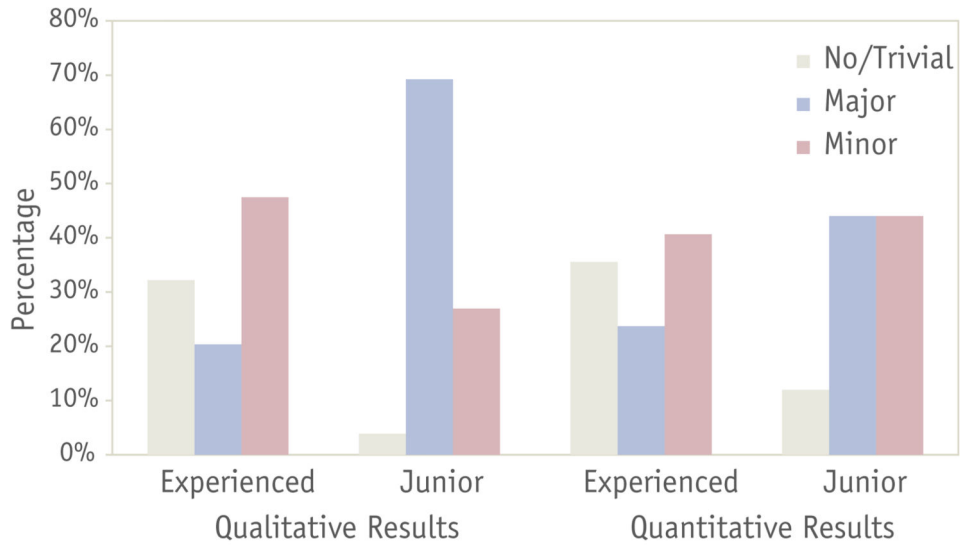


Fig. 2. Percentage for each type of qualitative and quantitative changes per physician experience. Quantitative change type was determined according to Dice similarity coefficient classification.

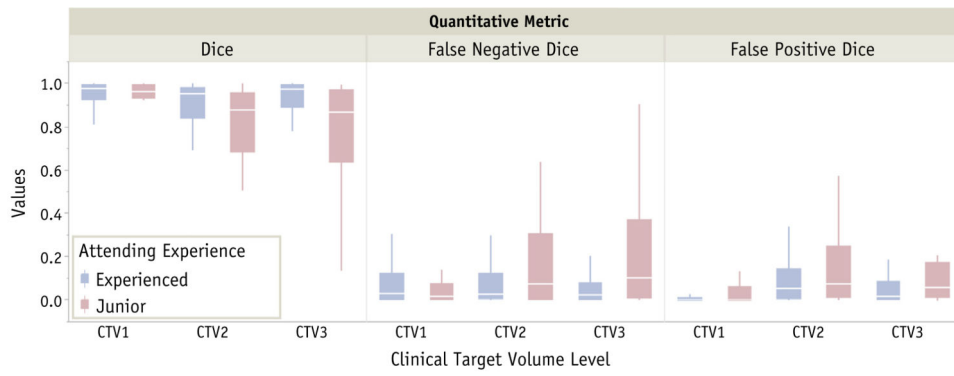


Fig. 3. Box plots of metrics used in the analysis for each target volume level per physician experience.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1
Patient and disease characteristics

Variables	n (%)
Age, y	
<40	4 (4.7)
40-65	50 (58.8)
>65	31 (36.5)
Sex	
Female	23 (27.1)
Male	62 (72.9)
Primary Site	
Oropharynx	27 (31.8)
Oral Cavity	20 (23.5)
Larynx	10 (11.8)
Skin	8 (9.4)
Salivary	6 (7.1)
Sinonasal	5 (5.9)
Unknown Primary	5 (5.9)
Thyroid	3 (4.0)
Nasopharynx	2 (2.4)
T stage	
Tx	2 (2.3)
T0	4 (4.7)
T1	11 (13)
T2	19 (22)
T3	14 (16.5)
T4	22 (26)
Recurrent	14 (16.5)
N stage	
Nx	1 (1)
N0	27 (32)
N1	14 (16.5)
N2	25 (29)
N3	4 (5)
Recurrent	14 (16.5)
Overall Staging	
Stage I	4 (4.7)
Stage II	4 (4.7)
Stage III	16 (18.8)
Stage IV	47 (55.3)
Recurrent	14 (16.5)
Treatment Intent	

Variables	n (%)
Adjuvant	48 (56.5)
Definitive	37 (43.5)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2
Quantitative change evaluation metrics per clinical target volume levels

CTV level	All Physicians (n=84)	Experienced (n=59)	Junior (n=25)	P Value
CTV1				
VD	-0.05 ± 0.13	-0.06 ± 0.10	-0.03 ± 0.18	.155
DSC	0.94 ± 0.09	0.95 ± 0.07	0.93 ± 0.13	.825
FND	0.09 ± 0.16	0.09 ± 0.12	0.10 ± 0.23	.668
FPD	0.03 ± 0.07	0.02 ± 0.05	0.05 ± 0.11	.932
CTV2				
VD	0.39 ± 2.91	0.05 ± 0.29	1.15 ± 5.21	.613
DSC	0.84 ± 0.23	0.87 ± 0.19	0.76 ± 0.29	.041
FND	0.14 ± 0.24	0.12 ± 0.20	0.21 ± 0.31	.320
FPD	0.18 ± 0.33	0.14 ± 0.23	0.27 ± 0.49	.401
CTV3				
VD	-0.02 ± 0.32	0.01 ± 0.28	-0.07 ± 0.39	1.000
DSC	0.87 ± 0.20	0.90 ± 0.18	0.80 ± 0.22	.006
FND	0.17 ± 0.34	0.11 ± 0.27	0.28 ± 0.43	.029
FPD	0.10 ± 0.17	0.09 ± 0.18	0.12 ± 0.17	.093

Abbreviations: CTV = clinical target volume; DSC = Dice similarity coefficient; FND = false negative Dice; FPD = false positive Dice; VD = volumetric difference.

Data represent means ± standard deviations.

P values were calculated with the Wilcoxon rank sum test to assess differences between experienced and junior physician contours.

Table 3
Quantitative classification from changes in clinical target volume contours

All Physicians N=84			
CTV level	Trivial, n (%)	Major, n (%)	Minor, n (%)
CTV1			
DSC	55 (65)	4 (5)	25 (30)
FND	49 (58)	13 (16)	22 (26)
FPD	67 (80)	2 (2)	15 (18)
CTV2			
DSC	38 (47)	18 (22)	25 (31)
FND	42 (52)	18 (22)	21 (26)
FPD	37 (46)	19 (23)	25 (31)
CTV3			
DSC	34 (51)	13 (20)	19 (29)
FND	39 (59)	15 (23)	12 (18)
FPD	39 (59)	10 (15)	17 (26)
Overall			
DSC	25 (30)	23 (27)	36 (43)
FND	25 (30)	35 (42)	24 (28)
FPD	27 (32)	30 (36)	27 (32)

Abbreviations: CTV = clinical target volume; DSC = Dice similarity coefficient; FND = false negative Dice; FPD = false positive Dice.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript