Practical Radiation Oncology® (2023) 13, e354−e364

practical radiation oncology
pro
www.practicalradonc.org

Basic Original Report

Effect of Radiation Therapy Quality Assurance on Nasopharyngeal Carcinoma: Usage of a Novel, Web-Based Quality Assurance Application

Jun Won Kim, MD, PhD,[a,b,1] Joseph Marsilla, BSc,[b,1] Michal Kazmierski, MSc,[b] Denis Tkachuk, BSc,[b] Shao Hui Huang, MD,[c] Wei Xu, PhD,[d] John Cho, MD,[c] Jolie Ringash, MD,[c] Scott Bratman, MD, PhD,[c] Benjamin Haibe-Kains, PhD,[b,e,*] and Andrew Hope, MD[c,*]

[a]*Department of Radiation Oncology, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Korea;* [b]*Princess Margaret Cancer Center, University Health Network, Toronto, Ontario, Canada;* [c]*Departments of Radiation Oncology;* [d]*Departments of Biostatistics, Princess Margaret Cancer Center/University of Toronto, Toronto, Ontario, Canada; and* [e]*Department of Computer Science, University of Toronto, Toronto, Ontario, Canada*

Received 21 September 2022; accepted 6 March 2023

Purpose: We used a new web application for rapid review of radiation therapy (RT) target volumes to evaluate the relationship between target delineation compliance with the international guidelines and outcomes of definitive RT for nasopharyngeal carcinoma (NPC).

Methods and Materials: The data set consisted of computed tomography simulation scans, RT structures, and clinical data of 354 patients with pathology-confirmed NPC treated with intensity modulated RT between 2005 and 2017. Target volumes were peer-reviewed in RT quality assurance rounds, and target contours were revised, if recommended, before treatment. We imported the contours of intermediate-risk clinical target volumes of the primary tumor (CTVp) of 332 patients into the application. Inclusion of anatomic sites within intermediate-risk CTVp was determined in accordance with 2018 international guidelines for CTV delineation for NPC and correlated with time to local failure (TTLF) using Cox regression.

Results: In the peer-review quality assurance analysis, local and distant control and overall survival rates were similar between peer-reviewed and nonreviewed cases and between cases with and without target contour changes. In the CTV compliance analysis, with a median follow-up of 5.6 years, 5-year TTLF and overall survival rates were 93.1% and 85.9%, respectively. The most frequently non-guideline-compliant anatomic sites were sphenoid sinus (n = 69, 20.8%), followed by cavernous sinus (n = 38, 19.3%), left and right petrous apices (n = 37 and 32, 11.1% and 9.6%), and clivus (n = 14, 4.2%). Among 23 patients with a local failure (6.9%), the number of noncompliant cases was 8 for sphenoid sinus, 7 cavernous sinus, 4 left and 3 right petrous apices, and 2 clivus. Cavernous sinus-

Sources of support: This work was supported by Canadian Institutes of Health Research Project #426366 and a National Research Foundation of Korea grant funded by the Korean Government Ministry of Science and Information & Communication Technology (NRF-2020R1A2C1102559).

Disclosures: none.

The web-based application, QUANNOTAE, is open-source under license Apache License 2.0 (github.com/bhklab/quannotate) and is freely accessible via https://www.quannotate.com. All the preprocessing scripts used to build the application are available at https://github.com/bhklab/QANPCRT. Data and statistical analyses are made fully reproducible through a published Code Ocean capsule (codeocean.com/capsule/1953079/tree). The radiologic data are publicly available on TCIA (submission process is ongoing).

* Corresponding authors: Benjamin Haibe-Kains, PhD; E-mail: benjamin.haibe.kains@utoronto.ca and Andrew Hope, MD; E-mail: andrew.hope@rmp.uhn.ca

[1] J.W.K. and J.M. contributed equally to this work.

https://doi.org/10.1016/j.prro.2023.03.003

conforming cases showed higher TTLF in comparison with nonconforming cases (93.6% vs 89.1%, $P = .013$). Multivariable analysis confirmed that cavernous sinus noncompliance was prognostic for TTLF.

Conclusions: Our application allowed rapid quantitative review of CTVp in a large NPC cohort. Although compliance with the international guidelines was high, undercoverage of the cavernous sinus was correlated with TTLF.

Introduction

Intensity modulated radiation therapy (IMRT) is a current standard option for the treatment of head and neck cancer (HNC) and often allows effective sparing of organs at risk (OARs). Dose-painting via IMRT allows delivery of nonuniform doses to clinical target volumes (CTVs) of different recurrence risks, thus improving the therapeutic ratio.[1] In the phase 2 Radiation Therapy Oncology Group 0225 trial of IMRT with or without chemotherapy for nasopharyngeal carcinoma (NPC), intermediate-risk CTV (IR-CTV) was defined as gross tumor volume (GTV) + 10-mm margin plus areas at risk for microscopic involvement. The recommended coverage for microscopic involvement was specific and included the entire nasopharynx, retropharyngeal nodal regions, skull base, clivus, pterygoid fossae, parapharyngeal space, sphenoid sinus, the posterior third of the nasal cavity/maxillary sinuses, and the pterygopalatine fossae.[2] The extent of IR-CTV for NPC is related to marginal local recurrences[3] and development of radiation toxicity, including xerostomia and swallowing difficulty,[4] yet interobserver variation in delineating IR-CTV can be substantial.[5] With advancement of conformal radiation therapy (RT) techniques and reduction of target margins, adequate target coverage has become a critical issue, and consensus guidelines for delineating neck node levels[6] and CTVs for conformal RT for nasopharynx and other head and neck tumor sites have been published.[7,8]

RT quality assurance (QA) is particularly important for HNC treatment, especially in the conformal RT era, due to the complexity of RT target volumes and multiple adjacent OARs.[9] A retrospective evaluation of protocol compliance in a phase 3 HNC RT trial found that protocol noncompliance was associated with a 20% poorer 2-year overall survival (OS), and 25% of major protocol violations were due to incorrect coverage of target volumes.[10] Clinical studies rely on assumption and/or maintenance of adequate QA; however, limitations in time and human resources to review large patient cohorts often result in compliance issues. Multi-institutional, prospective RT trials require central review of RT target delineation, yet only the first few cases are submitted and reviewed for participation in many trials.[11] Patterns-of-RT-practice studies often rely on survey questionnaires[12] instead of visually inspecting the RT structures, target volumes, and OARs.

To address these issues, we first reviewed our peer-reviewed RT-QA data and determined whether the routine clinical QA round was associated with a difference in treatment outcomes. Second, we used a cloud-based QA tool (*QUANNOTATE*) to independently evaluate the effect of RT target quality on treatment outcomes. *QUANNOTATE* allows rapid and qualitative review of RT target volumes in an easily accessible format without requiring access to the RT planning system (www.quannotate.com). Using this tool, we evaluated the relationship between target delineation compliance with the published international guidelines[7] and treatment outcomes of patients with NPC undergoing radical RT.

Methods and Materials

Study population and RT QA data analysis

The data set used for this study consisted of 354 patients with pathologically confirmed NPC who underwent IMRT at a tertiary referral cancer center between July 2005 and August 2017. All of these patients were included in an institutional, prospective HNC database in which outcomes are recorded at the point of care[13] and had been subject to the standard internal QA processes used at our center.[14] All patients were managed with multidisciplinary input according to institutional protocols. Generally, T1-T2 N0 patients received RT alone while T1-T2 N1 were treated with concurrent chemo-RT (if fit). More advanced cases were treated with concurrent chemo-RT followed by adjuvant chemotherapy. The RT regimen comprised 70 Gy in 35 fractions over 7 weeks using IMRT with daily image guidance.[15] RT target volumes were peer-reviewed in a weekly RT QA round, and target contours were revised, if recommended, before RT planning and delivery. Treatment outcome was analyzed with regards to whether RT target volumes were peer-reviewed, and changes in the target contours were made.

Cloud-based web application development

Among the 354 patients included in the RT QA data analysis, digital imaging and communications in medicine (DICOM) or RT structure data of 24 patients were inaccessible, leaving 332 patients available for the development of our cloud-based web application (Fig. 1). The data set consisted of anonymized computed tomography (CT) simulation scans, RT structures, and the clinical data. Patients' CT images contained a median of 181 slices, with each slice consisting of 512 × 512 pixels. The median slice thickness

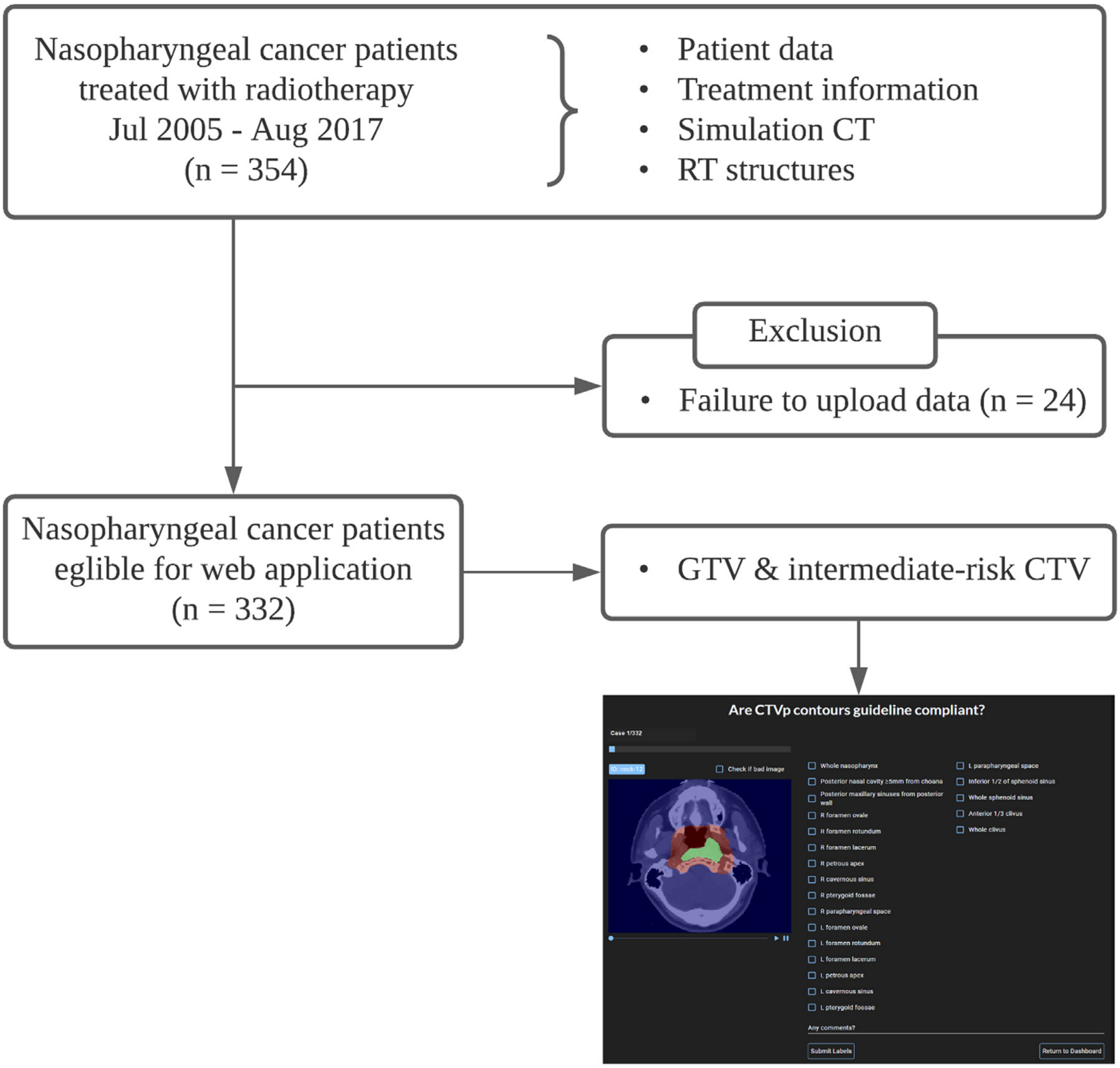

Figure 1 Study inclusion workflow.

was 2.0 mm (range, 2.0-3.0 mm).[16] The contours of GTV and IR-CTV were extracted from existing RT structure sets of each patient, and a python-based framework was used to develop the web application (Fig. E1). Details about implementation are provided in Supplementary Methods. This study conformed to the ethical guidelines of the 1975 Declaration of Helsinki and was approved by the Health Canada and Public Health Agency of Canada research ethics board (approval #17-5871).

IR-CTVp delineation compliance assessment

Two CTVs for the primary tumor (CTVp) are routinely used for NPC treatment at our institution: High risk CTV (HR-CTVp) was designated as CTV70 and IR-CTVp as CTV56, and these targets correspond to CTVp1 and CTVp2, respectively, in the international guidelines for the delineation of the CTV for NPC.[7] We loaded GTVp and IR-CTVp into our web application and used binary assessments of whether or not 18 anatomic sites were included within IR-CTVp: whole nasopharynx; posterior nasal cavity ≥5 mm from choana; posterior maxillary sinuses from the posterior wall; right and left foramen ovale, foramen rotundum, foramen lacerum, petrous apices, pterygoid fossae, and parapharyngeal spaces; ipsilateral cavernous sinus (if T3-4), inferior 1/2 (if T1-2) or whole (if T3-4) sphenoid sinus; and anterior 1/3 (if no invasion) or whole clivus (if invasion). Compliance score was defined as the number of anatomic sites included within IR-CTVp, and a score of 18 would indicate complete inclusion of these anatomic sites. The cloud-based web application can be accessed online from remote locations. Users are able to slide through an entire CT scan of a patient and review the CTV contours superimposed to the CT images. Blinded to the treatment outcomes, a radiation oncologist (JK) with 10 years of posttraining experience determined inclusion of the anatomic sites within IR-CTVp, and the results were exported to be analyzed after complete assessment.

Interpretation of international guidelines

Controversial issues may arise during the implementation of international guidelines in defining target volumes. For example, the expert group including key oncologists from Asia (China, Hong Kong, Korea, Singapore, Taiwan), Australia, North America (Canada, United States), Saudi Arabia, and Europe (Belgium, Denmark, France, The Netherlands, Turkey, United Kingdom) showed low (65%) consensus on whether or not to shave out air cavities within the HR-/IR-CTVp volumes for NPC.[7] The guidelines recommend coverage of the ipsilateral cavernous sinus for T3-4 tumors, yet some advanced tumors may not have clear invasion of the cavernous sinus or definite lateralization. Coverage of the petrous tip is recommended for all stages, yet the extent of adequate coverage for the petrous apex is not clear. We defined additional criteria to help determine guideline compliance: shaving of air cavity was not allowed for sphenoid sinus but for nasal cavity and maxillary sinuses; both cavernous sinuses should be covered in case invasion or laterality is unclear for T3-4 tumors, unless T4-category is solely related to inferior extension such as to the hypopharynx; and at least anterior 1/2 of the anterior petrous apex bounded by internal auditory canal should be covered in all stages (Table 1).

Statistical analysis

The Kaplan-Meier method was used to estimate OS and time to local failure (TTLF), and the log-rank test was used to provide a statistical comparison of 2 groups. Survival was calculated from the start of RT to death due to any cause or last patient follow-up and TTLF from the start of RT to a pathologically confirmed local failure or clear radiographic progression in cases where biopsy was not indicated. Factors with P values <.10 in univariable analysis (UVA) were evaluated for independence with the Cox proportional hazards model. We analyzed treatment outcomes of the whole cohort, whereas a subgroup T3-4 cases were analyzed when guidelines provided separate recommendations for locally advanced diseases. Data analysis was performed with SPSS Statistics, version 25.0 (IBM, Armonk, NY).

Results

We reviewed our RT-QA data, which was conducted to determine whether the routine clinical QA round was associated with a difference in treatment outcomes. Subsequently, we developed a new, cloud-based QA tool that allowed rapid and qualitative review of RT target volumes and evaluated the relationship between target delineation compliance with the published international guidelines and treatment outcomes of patients with NPC undergoing radical RT. The study workflow and the development process of the cloud-based web application are presented in Figs. 1 and E1, respectively.

Effect of RT QA rounds on treatment outcomes

Among the 354 patients, RT target volumes of 210 patients were peer-reviewed at weekly QA rounds, and the clinical characteristics of peer-reviewed and nonreviewed patients are compared in Table E1. Peer-reviewed patients were associated with shorter smoking history (P = .016) and a higher proportion of nondrinkers (P = .012) and showed a tendency toward higher (≥IVA-B) clinical stages (P = .068). The 5-year local control (LC; 91% vs 95%, P = .101), distant metastasis (DC; 85% vs 86%, P = .586), and OS (85% vs 85%, P = .946) rates were similar between the peer-reviewed and nonreviewed cases (Fig. E2A-C). Among the 210 peer-reviewed cases, RT target contours of 55 cases were revised as recommended by peer reviewers. Although the 5-year LC rates (90% vs 91%, P = .813) were similar between the cases with and without target contour changes, DC (79% vs 87%, P = .068) and OS rates (78% vs 87%, P = .069) were lower for the cases with changes in the target contours, although the difference was not statistically significant (P > .05; Fig. E2D-F).

Compliance of IR-CTVp with international guidelines

Clinical characteristics of 332 patients with NPC are summarized in Table 2. The median age was 52 years. T3-4 tumors were diagnosed in 59.3% and cervical lymph node metastasis in 87.3% of the patients. Concurrent chemoradiation was the predominant treatment modality (87.7%) with a median total RT dose of 70.0 Gy (Table 2). For all patients (n = 332), the number of cases with each anatomic site missed by IR-CTVp were as follows: 1 (0.3%) for whole nasopharynx, 2 (0.6%) for posterior nasal cavity, 2 (0.6%) for posterior maxillary sinuses, 6 (1.8%) and 4 (1.2%) for right and left foramen ovale, 14 (4.2%) and 12 (3.6%) for right and left foramen rotundum, 1 (0.3%) and 3 (0.9%) for right and left foramen lacerum, 0 and 1 (0.3%) for right and left pterygoid fossa, 0 and 1 (0.3%) for right and left parapharyngeal space, and 14 (4.2%) for the clivus. Right and left apices of the petrous bone were inadequately contoured in 32 (9.2%) and 37 (11.1%) cases, respectively, while the majority of inadequate coverage was observed in T1-2 cases. Inadequate contouring of sphenoid sinus was observed in 59

Table 1 Criteria for intermediate-risk CTVp compliance with international guidelines

Anatomic landmarks	International guidelines*	Additional criteria for guideline compliance
Margin from GTV	GTVp + 10 mm + whole NP	None
Nasal cavity, posterior part	At least 5 mm from choana	Uppermost cavity may be spared; shaving of air cavity is allowed
Maxillary sinuses, posterior part	At least 5 mm from posterior wall	Uppermost sinus may be spared; shaving of air cavity is allowed
Posterior ethmoid sinus	Include vomer	None
Skull base	Cover foramina ovale, rotundum, lacerum, and petrous tip	Foramina: Full coverage Petrous tip: At least anterior 1/2 of the anterior petrous apex bounded by internal auditory canal
Cavernous sinus	If T3-4 (involved side only)	Uppermost sinus may be spared; at least 5-mm margin superiorly from GTVp must be observed except for tumor abutting chiasm No definite laterality: both cavernous sinuses
Pterygoid fossae	+	None
Parapharyngeal spaces	Full coverage	None
Sphenoid sinus	Inferior 1/2 if T1-2; whole if T3-4	Air cavity should be included, but the uppermost sinus may be spared for T3-4
Clivus	1/3 if no invasion; whole if invasion	Upper- and lowermost part may be spared, but at least 5-mm margin from invasion front must be observed except for tumor abutting brain stem
Minimal margin if tumor in close proximity to critical OARs	GTVp + 2 mm	None

Abbreviations: CTVp = clinical target volume of the primary tumor; GTVp = gross tumor volume of the primary tumor; NP = nasopharynx; OARs = organs at risk.

* The international guidelines are considered minimum requirements for guideline compliance in the present study.

(29.9%) of the T3-4 cases and 10 (7.4%) of the T1-2 cases. Contouring of cavernous sinus was recommended only for T3-4 tumors, and 38 of the 197 cases (19.3%) were inadequately covered (Fig. 2A).

Patterns of failure

At a median follow-up of 5.6 years (range, 0.1-13.8), treatment resulted in 13 (3.9%) local, 3 (0.9%) regional, 37 distant (11.1%), 7 (2.1%) locoregional, 2 (0.6%) local and distant, 9 (2.7%) regional and distant, and 1 (0.3%) locoregional and distant failures (Fig. E3). The ratio of nonguideline contoured anatomic sites among 23 local failures to those among all patients was highest for left (1/4) and right (1/6) foramen ovale, followed by cavernous sinus (7/38), clivus (2/14), sphenoid sinus (8/69), left (4/37) and right (3/32) petrous apices, and left (1/12) and right (1/14) foramen rotundum (Fig. 2B). The compliance score and number of local failures for each of the scores are shown in Fig. E4. Score of 16 or higher was achieved in 93.9% of the cases, and local failure rates increased from 5.4% for a score of 18 to 20.0% for a score of 15.

Survival analysis

T4 disease (P = .006) and undercontouring of the cavernous sinus (P = .013) showed a significant correlation with decreased TTLF, and treatment with RT alone (P = .084) showed a marginal correlation in UVA, while these variables all proved to be independent prognostic factors in multivariable analysis (MVA) (Table 3). A similar trend was observed in a subgroup of T3-4 cases (n = 197), albeit cavernous sinus undercontouring (P = .072) was less prognostic for TTLF than T4 (P = .008) and RT alone (P = .049) in MVA (Table E2). Nodal staging (N0-2 vs N3) was included in the analyses of recurrence-free survival

Table 2 **Patient characteristics**

Variable Age		Number of patients Median, 52 y (range, 16-88)	%
Sex	Male	238	71.7
	Female	94	28.3
Performance	ECOG 0	236	71.1
	ECOG 1	89	26.8
	ECOG 2	3	0.9
	ECOG 3	2	0.6
	Unknown	2	0.6
Smoking history	Current	55	16.6
	Ex-smoker	79	23.8
	Nonsmoker	189	56.9
	Unknown	9	2.7
Drinking history	Ex-drinker	14	4.2
	Heavy (>20 drink/wk)	11	3.3
	Moderate (10-20 drink/wk)	8	2.4
	Light (<10 drink/wk)	53	16.0
	Nondrinker	230	69.3
	Unknown	16	4.8
Epstein-Barr virus	Positive	241	72.6
	Negative	20	6.0
	No test	71	21.4
Pathology	Type 1 (WHO I)	10	3.0
	Type 2 (WHO IIA)	54	16.3
	Type 3 (WHO IIB)	266	80.1
	Neuroendocrine (small cell)	2	0.6
T stage	T1	98	29.5
	T2	37	11.1
	T3	105	31.6
	T4	92	27.7
N stage	N0	42	12.7
	N1	79	23.8
	N2	164	49.4
	N3	47	14.1
Overall stage	I	14	4.2
	II	39	11.7
	III	151	45.5
	IVA	81	24.4
	IVB	47	14.2
Treatment	CCRT	291	87.7
	RT alone	41	12.3
Total RT dose		Median, 70.0 Gy (range, 59.4-70.0)	
RT fractions		Median, 35 fractions (range, 25-35)	

Abbreviations: CCRT = concurrent chemoradiation; ECOG = Eastern Cooperative Oncology Group; RT = radiation therapy; WHO = World Health Organization.

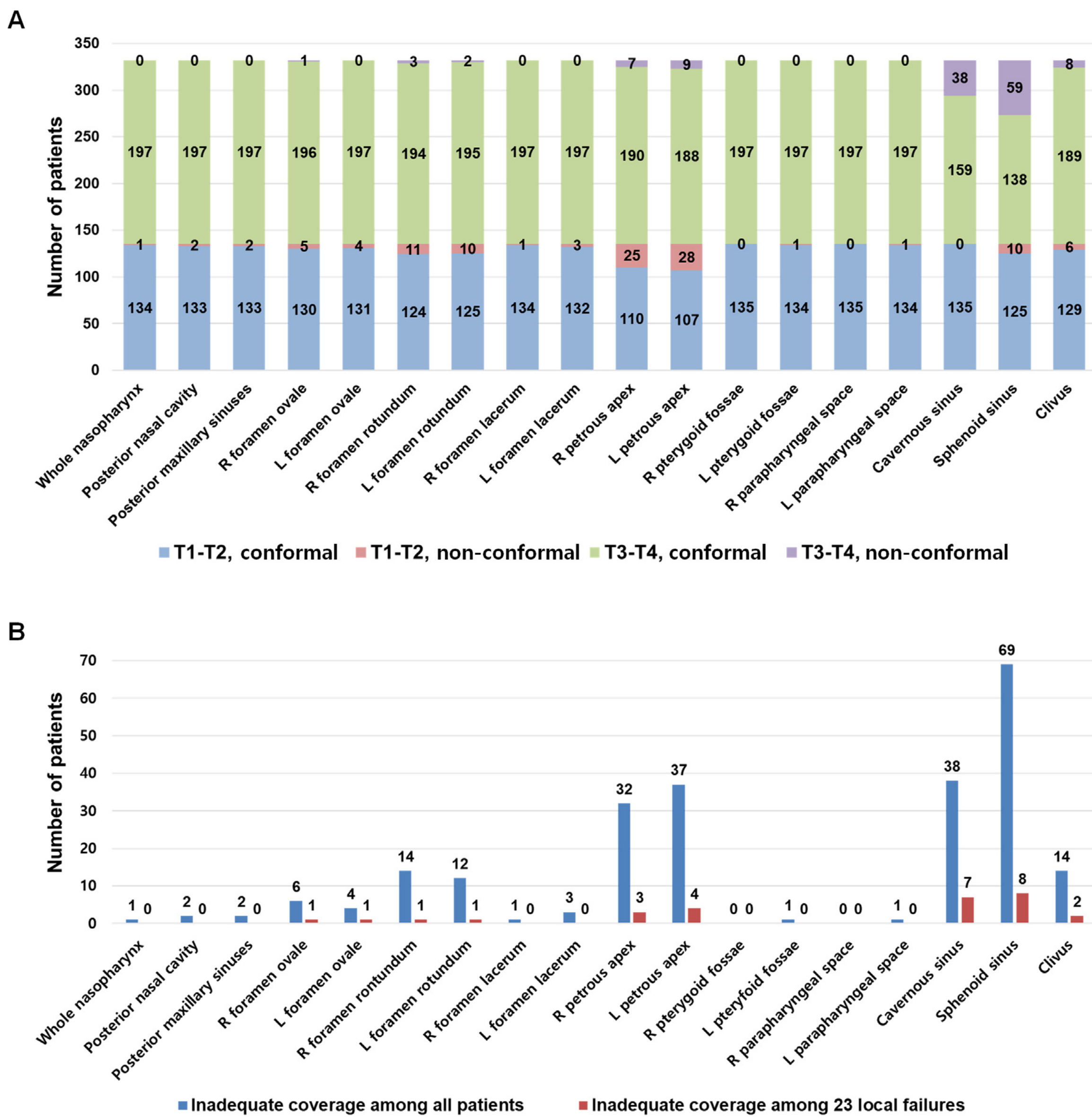

Figure 2 Coverage of anatomic sites by intermediate-risk clinical target volume. (A) Compliant versus noncompliant contours among patients with T1-2 and T3-4 tumors. (B) Noncompliant contours among all patients versus 23 local failures.

and OS, and the results of UVA and MVA are summarized in Tables E3 and E4, respectively. T4 (P = .002) and N3 (P < .0001) were independently prognostic for recurrence-free survival, whereas age >53 (P = .023), Eastern Cooperative Oncology Group ≥1 (P = .003), World Health Organization I/IIa (P = .048), T4 (P = .001), N3 (P < .0001), and RT alone (P = .003) showed statistically significant correlation with poor OS in MVA. Five-year TTLF and OS rates were 93.1% and 85.9% for all patients and 91.5% and 80.8% for the T3-4 subgroup (Fig. E5). For cavernous sinus-conforming versus nonconforming to the guidelines, 5-year TTLF showed significant differences (93.6% vs 89.1%, P = .013) while OS did not (87.8% vs 74.0%, P = .099; Fig. 3).

Discussion

In radiation oncology, peer review of target volumes and treatment plans ensures quality of radiation delivery and thus improved treatment outcomes. In our internal RT-QA analysis, LC, DC, and OS rates were similar whether or not the case was reviewed by weekly QA rounds. However, similar LC was achieved by the patients whose target contours were revised compared with those without contour changes, despite the fact that DC and OS of patients with revised contours indicate a more aggressive nature of the disease. Limitations of the peer review process in QA rounds include subjective and qualitative nature of these reviews. The overall compliance using the cloud-based

Table 3 Prognostic factors for time to local failure (n = 332)

Variable		No.	Univariate analysis			Multivariate analysis		
			HR	95% CI	*P*	HR	95% CI	*P*
Age	>53	157	2.00	0.87-4.64	.10			
	≤53	175	1					
Sex	Female	94	1.59	0.69-3.67	.290			
	Male	238	1					
Performance	ECOG ≥1	96	1.61	0.68-3.81	.28			
	ECOG 0	236	1					
Smoking Hx	Yes/unknown	143	0.78	0.33-1.84	.57			
	None	189	1					
Alcohol Hx	Yes/unknown	102	1.1	0.43-2.81	.84			
	None	230	1					
EBV	Negative	91	0.81	0.31-2.10	.669			
	Positive	241	1					
Pathology	WHO I/IIa	66	0.96	0.33-2.78	.937			
	WHO IIb	266	1					
T stage	T4	92	3.15	1.39-7.14	.006	4.02	1.63-9.94	.003
	≤T3	240	1			1		
Treatment	RT alone	41	2.4	0.89-6.48	.084	4.50	1.50-13.5	.007
	CCRT	291	1			1		
Compliance	<18	127	1.91	0.84-4.35	.13			
	18	205	1					
Petrous apex	Nonconforming	48	1.51	0.56-4.1	.41			
	Conforming	284	1					
Cavernous sinus	Nonconforming	38	3.07	1.26-7.48	.013	2.92	1.19-7.19	.020
	Conforming	294	1			1		
Sphenoid sinus	Nonconforming	69	2.04	0.87-4.82	.10			
	Conforming	263	1					

Abbreviations: CCRT = concurrent chemoradiation; CI = confidence interval; EBV = Epstein-Barr virus; ECOG = Eastern Cooperative Oncology Group; HR = hazard ratio; Hx = history; RT = radiation therapy; WHO = World Health Organization.

application was not associated with improved outcomes; however, undercovering the cavernous sinus was associated with higher local failure on UVA and MVA.

Multi-institutional clinical trials involving modern RT techniques are unique in that treatment quality can be assessed by reviewing digitized medical images and associated RT structures.[17,18] Ng et al[19] showed that when the volume within GTVp <66.5 Gy (95% prescribed dose) was greater than the cutoff volume of 3.4 cm^3, the 5-year local failure-free survival decreased from 90% to 54%. This suggests that survival, as function of target coverage, is highly nonlinear, where even a small "cold spot" (ie, 3.4 cm^3 implies sphere of radius 0.93 cm) can have an enormous effect (ie, 40% reduction in survival from 90%-54%). MVA by Ng et al showed that underdosing was an independent predictor of local failure, disease free survival, and OS. In the modern conformal RT era, QA of large-scale trials is challenging because of time-consuming processes for collecting and reviewing hundreds or thousands of individual cases and difficulty in eliminating interobserver variability among reviewers. We developed *QUANNOTATE*, a cloud-based web application for rapid review of hundreds or thousands of NPC cases (and other primary sites). *QUANNOTATE* provides a QA platform for uploading CT simulation images and RT structures and allows rapid and qualitative review of target volumes with respect to published guidelines in an easily accessible format.

With a median follow-up of 5.6 years in the current study, the local recurrence rate of 6.9% for NPC is comparable with previously published IMRT data.[20,21]

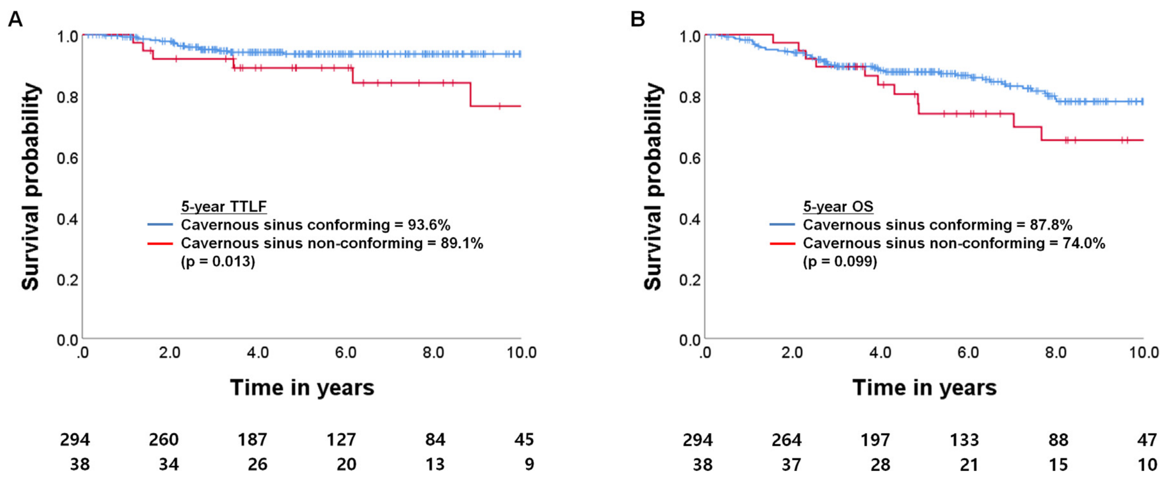

Figure 3 Kaplan-Meier curves for (A) time to local failure and (B) overall survival by cavernous sinus conformity.

Nevertheless, inadequate coverage of the cavernous sinus was observed in 19.3% of T3-4 cases and was an independent prognostic factor for TTLF. The importance of including the cavernous sinus within IR-CTVp was not universally recognized: Among the 6 studies reviewed by the international guidelines committee,[7] coverage of the cavernous sinus was not stated in 2 studies, Radiation Therapy Oncology Group 0225[2] and the Chinese guidelines.[22] The recommendation to include the involved-side cavernous sinus for T3-4 tumor was based on the findings by Liang et al,[23] where the cumulative incidence rate of cavernous sinus invasion was 17.4%; however, the structure was only at high risk when the tumor infiltrated the petrous apex or the foramen lacerum. Liang et al suggested that local disease tends to spread stepwise from proximal site to more distal sites, thus covering cavernous sinus is necessary for T3-4 tumors. Our data support the international guidelines' recommendations on the coverage of the cavernous sinus for T3-4 tumors. High local failure rates associated with inadequate coverage of other anatomic sites, including foramen ovale and sphenoid sinus, suggest that, with a larger study cohort, we may be able to "rank" these anatomic sites in order of importance in delineating target volumes.

Imaging biomarker studies are rapidly gaining popularity. Radiomics studies extract features from a large number of radiologic medical images using diverse computational approaches.[24-26] Although measures are taken to ensure reproducibility of the results via standardizing processes of image acquisition and data analysis,[27,28] the effect of treatment quality is often overlooked in radiomics studies, potentially confounding the search for robust biomarkers. With respect to optimizing plans, standardizing treatment to follow known, accepted international guidelines minimizes variance. These changes are "low hanging fruit" and, arguably, the simplest, most effective change one can do to improve treatment outcomes. If a chemotherapy agent demonstrated a comparable positive effect as rigorous treatment QA, it would have been adopted without question. Our results suggest that radiomics analysis of published data sets should be preceded by detailed QA analysis to ensure outcomes are not confounded because of variance in treatment-related factors as opposed to tumor factors. This tool could also be used as a measure for treatment quality but in a more general sense. For example, the proportion of plans that conform within the guidelines could be used as a quality metric at the level of the individual, site group, department, or even institution.

Our cloud-based application is not well suited for reviewing a small number of cases in clinical rounds, during which participants can easily access an RT planning system to review individual cases in detail. The design of our application in the current form is better suited for reviewing a large number of cases among multidisciplinary teams or multi-institutional trial groups. The presentation format of *QUANNOTATE* can be modified according to the objectives of its application, and we can suggest several areas where use of *QUANNOTATE* can be beneficial. First, the target delineation QA platform can readily be used for assessment of anatomic understanding and pathway of tumor spread and can be an effective tool for teaching and assessing medical students and radiation oncology residents. Second, RT planning studies often review and correlate radiation dose distribution with treatment outcomes. Dose distribution data stored in RT planning stations can readily be transported onto *QUANNOTATE* and reviewed with corresponding RT target volumes. Third, upgraded with additional viewing windows and a toggle switch for image selection, *QUANNOTATE* can be used to compare regions of interest between serially taken images, and this can be an effective tool for evaluating treatment responses in studies with large sample sizes. Our goal is to implement artificial intelligence-based image recognition in the *QUANNOTATE* platform, and this will reduce time consumption and interobserver variability and open doors to many more applications.

This study is subject to the limitations of all retrospective and single-institution studies, such as selection bias, small number of events, and observer bias. The results of the current study were derived from a largely endemic population that did not undergo induction chemotherapy. We focused on the compliance of IR-CTVp with the international guidelines, thus the outcome of interest was mainly local failure. The NPC cases imported into *QUANNOTATE* include the cases whose target contours had been revised per recommendation during internal QA rounds. With high compliance of target volumes and a small number of events (23 local failures), we lacked power to examine the significance of other anatomic sites. The analysis of local failure sites in relation to anatomic structures and radiation dose distribution is not included in the current study because of the close proximity of the anatomic structures reviewed for protocol compliance and lack of dose distribution data. In accordance with the study by Liang et al,[23] 1 of 4 to 6 (17%-25%) cases with inadequate coverage of foramen ovale resulted in local failures in the current study. A validation study with a large external cohort would increase the power to examine the effect of undercoverage of other anatomic sites. Second, compliance of IR-CTVp with the international guidelines was reviewed by only 1 experienced radiation oncologist. We implemented additional criteria of compliance to minimize observer bias by the reviewer. A validation study with an external cohort and interobserver variation analysis is underway.

Conclusion

We developed a web-based QA tool for qualitative review of target volumes for a large NPC cohort. Although target volumes of the study cohort showed high compliance with the international guidelines, *QUANNOTATE* identified inadequate coverage of the cavernous sinus as correlated with local failure, in agreement with the guidelines.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.prro.2023.03.003.

References

1. Gregoire V, Langendijk JA, Nuyts S. Advances in radiotherapy for head and neck cancer. *J Clin Oncol*. 2015;33:3277-3284.
2. Lee N, Harris J, Garden AS, et al. Intensity-modulated radiation therapy with or without chemotherapy for nasopharyngeal carcinoma: Radiation Therapy Oncology Group phase II trial 0225. *J Clin Oncol*. 2009;27:3684-3690.
3. Li JX, Huang SM, Jiang XH, et al. Local failure patterns for patients with nasopharyngeal carcinoma after intensity-modulated radiotherapy. *Radiat Oncol*. 2014;9:87.
4. Miao J, Di M, Chen B, et al. A prospective 10-year observational study of reduction of radiation therapy clinical target volume and dose in early-stage nasopharyngeal carcinoma. *Int J Radiat Oncol Biol Phys*. 2020;107:672-682.
5. Cardenas CE, Mohamed ASR, Tao R, et al. Prospective qualitative and quantitative analysis of real-time peer review quality assurance rounds incorporating direct physical examination for head and neck cancer radiation therapy. *Int J Radiat Oncol Biol Phys*. 2017;98:532-540.
6. Gregoire V, Ang K, Budach W, et al. Delineation of the neck node levels for head and neck tumors: A 2013 update. DAHANCA, EORTC, HKNPCSG, NCIC CTG, NCRI, RTOG, TROG consensus guidelines. *Radiother Oncol*. 2014;110:172-181.
7. Lee AW, Ng WT, Pan JJ, et al. International guideline for the delineation of the clinical target volumes (CTV) for nasopharyngeal carcinoma. *Radiother Oncol*. 2018;126:25-36.
8. Gregoire V, Evans M, Le QT, 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.
9. Van Gestel D, Dragan T, Gregoire V, Evans M, Budach V. Radiotherapy quality assurance for head and neck squamous cell carcinoma. *Front Oncol*. 2020;10:282.
10. 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.
11. Giaddui T, Yu J, Manfredi D, et al. Structures' validation profiles in Transmission of Imaging and Data (TRIAD) for automated National Clinical Trials Network (NCTN) clinical trial digital data quality assurance. *Pract Radiat Oncol*. 2016;6:331-333.
12. Sung SY, Kang MK, Kay CS, et al. Patterns of care for patients with nasopharyngeal carcinoma (KROG 11-06) in South Korea. *Radiat Oncol J*. 2015;33:188-197.
13. Wong K, Huang SH, O'Sullivan B, et al. Point-of-care outcome assessment in the cancer clinic: Audit of data quality. *Radiother Oncol*. 2010;95:339-343.
14. Huang SH, O'Sullivan B, Su J, et al. Hypofractionated radiotherapy alone with 2.4 Gy per fraction for head and neck cancer during the COVID-19 pandemic: The Princess Margaret experience and proposal. *Cancer*. 2020;126:3426-3437.
15. Chin O, Yu E, O'Sullivan B, et al. Prognostic importance of radiologic extranodal extension in nasopharyngeal carcinoma treated in a Canadian cohort. *Radiother Oncol*. 2021;165:94-102.
16. Arrowsmith C, Reiazi R, Welch ML, et al. Automated detection of dental artifacts for large-scale radiomic analysis in radiation oncology. *Phys Imaging Radiat Oncol*. 2021;18:41-47.
17. Fairchild A, Aird E, Fenton PA, et al. EORTC Radiation Oncology Group quality assurance platform: Establishment of a digital central review facility. *Radiother Oncol*. 2012;103:279-286.
18. Ohri N, Shen X, Dicker AP, Doyle LA, Harrison AS, Showalter TN. Radiotherapy protocol deviations and clinical outcomes: A meta-analysis of cooperative group clinical trials. *J Natl Cancer Inst*. 2013;105:387-393.
19. Ng WT, Lee MC, Chang AT, et al. The impact of dosimetric inadequacy on treatment outcome of nasopharyngeal carcinoma with IMRT. *Oral Oncol*. 2014;50:506-512.
20. Liu X, Tang LL, Du XJ, et al. Changes in disease failure risk of nasopharyngeal carcinoma over time: Analysis of 749 patients with long-term follow-up. *J Cancer*. 2017;8:455-459.
21. Tian YM, Liu MZ, Zeng L, et al. Long-term outcome and pattern of failure for patients with nasopharyngeal carcinoma treated

with intensity-modulated radiotherapy. *Head Neck*. 2019;41: 1246-1252.

22. 2010 Nasopharyngeal carcinoma intensity modulated radiotherapy target and dose design guidelines. Expert consensus. *Chin J Radiat Oncol*. 2011;20:267-269.
23. Liang SB, Sun Y, Liu LZ, et al. Extension of local disease in nasopharyngeal carcinoma detected by magnetic resonance imaging: improvement of clinical target volume delineation. *Int J Radiat Oncol Biol Phys*. 2009;75:742-750.
24. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. *Radiology*. 2016;278:563-577.
25. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48:441-446.
26. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018;18:500-510.
27. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14:749-762.
28. Zwanenburg A, Vallieres M, Abdalah MA, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295:328-338.