Advances in Radiation Oncology (2024) 9, 101591

advances in radiation oncology

www.advancesradonc.org

Scientific Article

Interobserver Variability in a Spanish Society of Radiation Oncology (SEOR) Head and Neck Course. Is Current Contouring Training Sufficient?

Victor De La Llana, MSc,[a,b] Fernando Mañeru, PhD,[a,b,*] Julián Librero, MSc,[c,d] Santiago Pellejero, MSc,[a,b] and Fernando Arias, MD, PhD[b,e]

[a]Department of Medical Physics, Hospital Universitario de Navarra (HUN), Pamplona, Spain; [b]Navarra Institute for Health Research (IdiSNA), Pamplona, Spain; [c]Navarrabiomed, Hospital Universitario de Navarra (HUN) − Universidad Pública de Navarra (UPNA), Pamplona, Spain; [d]Research Network on Chronicity, Primary Care and Health Promotion (RICAPPS), Madrid, Spain; and [e]Department of Radiation Oncology, Hospital Universitario de Navarra (HUN), Pamplona, Spain

Received 15 March 2024; accepted 26 July 2024

Purpose: External beam radiation therapy has grown significantly, incorporating advanced techniques like intensity modulation or stereotactic treatments, which enhance precision and accuracy. Nevertheless, variability in target volume delineation by radiation oncologists remains a challenge, influencing dose distribution. This study analyzes an online training course by the Spanish Society of Radiation Oncology, focusing on head and neck tumor contouring, to evaluate interobserver variability.

Material and Methods: Eight instructors provided clinical directives for 8 head and neck pathologies. Participants contoured structures using their own treatment planning systems, emphasizing gross tumor volume and high-, medium-, and low-risk clinical target volumes (CTV) contouring. Delineation variability was evaluated using the Dice similarity coefficient and volume relative change.

Results: The results reveal significant variability in contouring, with mean Dice similarity coefficient values ranging from 0.57 to 0.69. High-risk CTV demonstrated higher variability compared with medium-risk CTV. The presence of a gross tumor volume and supporting positron emission tomography/computed tomography or magnetic resonance imaging studies did not significantly improve the concordance. Parotid cases exhibited the greatest differences.

Conclusions: Despite the introduction of new automatic tools, this study points to the need for uniform contouring criteria. Training and standardization efforts are essential to enhance radiation therapy treatment consistency and quality.

Introduction

Computed tomography (CT) contouring for planning purposes needs anatomic knowledge and a comprehensive analysis of the clinical details of each case. This task is carried out by radiation therapy oncologists, specialized physicians with years of training.

Unfortunately, differences in structure delineation were found between different observers, although they are expert oncologists.[1,2] International consensus guidelines for delineation of the clinical target volumes (CTV) and organs at risk (OAR) in head and neck (H&N) cancers have also been published[3-6] to reduce these variations, but there is no total consensus. The quantification of these

Sources of support: This work had no specific funding.

Data Sharing Statement: Research data are stored in an institutional repository and will be shared on request to the corresponding author.

*Corresponding author: Fernando Mañeru, PhD; Email: fernando.maneru.camara@navarra.es

https://doi.org/10.1016/j.adro.2024.101591

differences has been a challenge for decades, and only limited data are available in the literature. The main uncertainty in radiation therapy treatment comes from these differences in contouring that result in variability in the dose−volume metrics.[7]

Recent technological advances have made the generalization of automated contouring aid tools in treatment planning systems possible. Artificial intelligence algorithms and deep learning methods open up new horizons in terms of the delimitation of structures. All of these advances are great improvements, but the final decision still requires careful contour review by a specialized oncologist. The unification of criteria and the standardization of the process are the only possible solutions in any technological context of segmentation.

An example of the latter is the practical contouring courses for oncologists of the Spanish Society of Radiation Oncology, whose latest edition focused on H&N tumors.

The objective of this work was to analyze the experience of this course and obtain conclusions about this current concern. A study of systematic qualitative differences in contouring and quantitative results is presented, using Dice similarity coefficient (DSC) and volume relative change (ΔV), which will be defined in the material and methods section. With the purpose of improving our knowledge of these potential variables analysis, we suggested 2 complementary hypotheses related to 2 features, a priori considered favorable factors, for reducing variability.

Complementary hypothesis 1: CTV high-risk delineation with and without the presence of gross tumor volume

One of the hypotheses of the study is that the presence of a gross tumor volume (GTV) in the patient, that is, nonoperated patient, increases the high-risk CTV (CTV_HR) degree of agreement between students and teachers. Our goal with this hypothesis was to verify that a clear area, which defines the real tumor volume (GTV), helps in the contouring of the next level (CTV_HR).

To analytically assess this hypothesis, a multiple linear regression model was applied with the study case as a random effect, the DSC as the dependent variable, and the presence or absence of GTV as a fixed explanatory variable.

Complementary hypothesis 2: supporting studies

The use of supporting studies such as positron emission tomography/CT (PET/CT) and MRI makes it easier to reduce the variability between different observers.[8,9] Deantonio et al[10] published that thanks to the use of PET/CT, it was possible to contour a larger GTV, and currently, PET/CT is a very useful tool in tumor staging, tumor control, and validation of the effect of radiation therapy. Instead, magnetic resonance imaging (MRI) emerged as a new promising image modality allowing for better tumor and soft tissue visualization, which can help targets and OAR contouring.

In this course, there are cases with supporting studies, and we hypothesize that with this assistance, contours would converge better to the reference (instructor delineation). A particular case was the parotid: students have access to a PET/CT acquired before surgery; hence, this pathology was considered as if it had no image in the complementary hypothesis 2 analysis.

Materials and Methods

Course description

The Spanish Society of Radiation Oncology organized a medical online training course ("Practicum") focusing on practical learning of contouring in H&N radiation therapy.

Participants were residents in training or junior physicians, all clinical professionals. Enrolment for the course was not mandatory in the residence program. Therefore, students had to pay a registration fee to attend. The exact data for training years of each one were not available in this study.

In 2021, 8 instructors, each with 1 different pathology, gave clinical directives on how to contour. The course was taught online due to the continuing limitations in place during the recent pandemic caused by the SARS-CoV-2 (COVID-19).

The trainees received a full description of the cases online. They imported the planning CT into their treatment planning system (TPS), and once the contouring was completed, they sent the cases to the course organizers.

Patient cases

Eight H&N cases were selected for this course: parotid, larynx, paranasal sinuses, oral cavity, oropharynx, nasopharynx, stereotactic body radiation therapy, and cervical metastases. All the data were anonymized before its use for the course and posterior investigation purposes. Participants consented to the use of their contours for post-course investigation.

All images were acquired in a supine position with a dedicated radiation therapy planning CT with a slice not superior to 3 mm. In some of the cases, supporting diagnostic imaging (MRI or PET/CT) was added for more

detailed information. PET/CT images were available for parotid, larynx, and oropharynx cases, whereas MRI images were available for paranasal sinuses and nasopharynx cases.

Students' delineation

Participants imported the 8 sets of DICOM (Digital Imaging and Communication in Medicine) objects in their own TPS, and structures can be delineated at their home centers. A 1-month installment was set.

It was encouraged to segment clinical targets and some OAR in each case. Targets were divided into 4 groups: GTV, CTV_HR, medium-risk CTV (CTV_MR), and low-risk CTV (CTV_LR).

In Fig. 1, we can see an example of the larynx CTV_HR contour for all the students who completed this pathology.

Twenty-three participants took part in the 2021 online course with the condition to contour 5 of the 8 cases to achieve the course.

Therefore, students did not have to complete the delineation of all cases, and they developed the entire contouring process prior to any expert solution.

Treatment prescriptions were established in case description, including the number of fractions, phases (integrated or not), dose fractionation, and total target doses. These pre-established conditions seek to focus the differences only on target contouring and eliminate as far as possible other possible factors that may affect the treatment decision.

Contouring variability study

For the analysis, students' contours were exported from their different TPSs and imported into the same TPS. Comparison was done in Eclipse v 16.0 (Varian MS).

Our reference structure was the one contoured by the instructor since the students were expected to follow his/her instructions for contouring.

Different metrics were studied to analyze interobserver variability and variances in contouring following the same clinical indications.

1. Volume relative change (ΔV): Subtraction of each participant contouring (V_i) and reference structure (V_R), instructor contouring, divided by V_R:

$$\Delta V(V_i, V_R) = \frac{V_i - V_R}{V_R}$$

2. DSC. This volume overlap metric is calculated by doubling the volume that overlaps between V_i and V_R and dividing it by the sum of the volumes (V_i, V_R). A DSC value equal to 1 means that V_i and V_R are the same size and shape. A DSC close to 0 means that there is minimal overlap between V_i and B or a large difference in the size ratio between them (Fig. 2).

$$DSC = \frac{2|Vi \cap V_R|}{|Vi| + |V_R|}$$

Global delineation compatibility was assessed with DSC records on a scale (0, 1) (1-DSC accounts for the degree of dissimilarity). Analyzing graphically their distribution, between organs and/or for the different targets, allows us to relate, for different pathologies and targets, the bias in the delineation (DSC average: using as reference the global mean of agreement) and the accuracy of the delineation (DSC variance: using as reference global mean of agreement variance).

To test the hypothesis that raises improvements in contouring if GTV is available, the degree of DSC achieved in relation to the presence of GTV was analyzed using multiple regression with pathology (n = 8) as a random effect.

Results

Number of contourings divided by pathology and target volume were as follows:

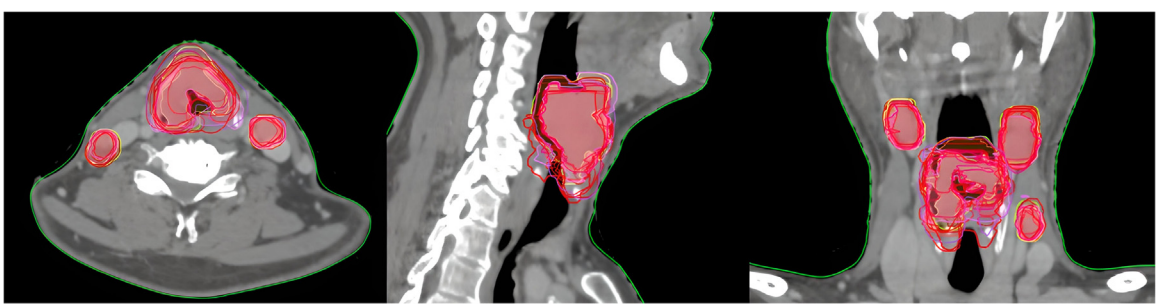

Figure 1 Larynx CTV_HR course contours.
Abbreviation: CTV_HR = high-risk clinical target volume.

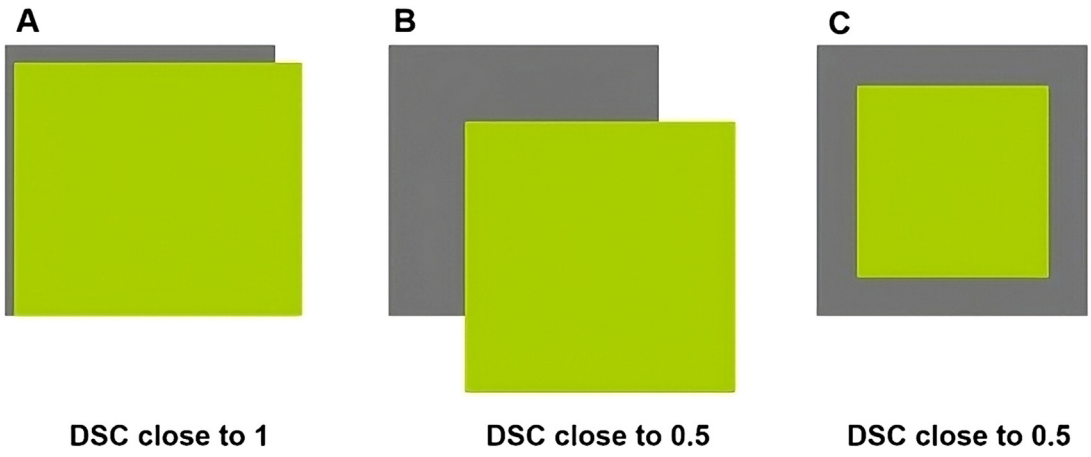

Figure 2 DSC concept: (A) good agreement; (B, C) different bad agreement examples with same DSC value.
Abbreviation: DSC = dice similarity coefficient.

- GTV: 8 parotid gland, 14 larynx, 19 oropharynx, 10 nasopharynx, and 16 stereotactic body radiation therapy.
- CTV_HR: 17 parotid gland, 14 larynx, 16 paranasal sinuses, 13 oral cavity, 19 oropharynx, 8 nasopharynx, and 12 cervical mets.
- CTV_MR: 17 parotid gland, 14 larynx, 16 paranasal sinuses, 13 oral cavity, 19 oropharynx, 8 nasopharynx, and 13 cervical mets.
- CTV_LR: 17 parotid gland and 14 larynx.

DSC analysis

Table 1 illustrates the statistical summary of data related to DSC detailed by target type. There were a minimum number of 10 delineations per pathology (nasopharynx) and a maximum of 20 (oropharynx).

Mean DSC and standard deviation (SD) for GTV, CTV_HR, CTV_MR, and CTV_LR were 0.57 ± 0.16, 0.57 ± 0.13, 0.66 ± 0.09, 0.69 ± 0.05, respectively. Global DSC mean and SD values obtained were 0.61 ± 0.12. Coefficient of variation (CV) mean for all pathologies was 0.21 (0.05; 0.65).

An example of CTV_HR variability contouring divided by pathologies is shown in the box plot graph (Fig. 3). CTV_HR is one of the most important targets as it is prescribed to the highest dose (the same as GTV) and is delineated straight from GTV without taking into consideration anatomic references.

Figure 3 suggests that pathologies without a GTV present have a similar pattern with respect to the CTV_HR (complementary hypothesis 1 false). A quantitative evaluation of this hypothesis found that the presence of GTV, analyzed in the set of pathologies, does not change the average of the DSC parameter (beta 0.01 between −0.14 and +0.16).

In Fig. 4A, we can see a detailed analysis taking into account pathologies for CTV_HR. The horizontal and vertical solid lines across Fig. 4A represent the global DSC mean value and global SD mean value. Paranasal sinuses and larynx targets show points above the global DSC mean line representing a better agreement between students and instructors. Moreover, these points are located under global SD mean line showing a low dispersion among all contours. Parotid gland shows the lowest mean DSC and the highest SD. Larynx shows the opposite situation, with the highest mean DSC and the lowest SD. Both the Parotid gland and the Oropharynx SD are worse than the global SD mean value.

DSC detailed also by target is presented in Fig. 4B. CTV_MR has a high DSC and little dispersion, regardless of pathology, whereas CTV_HR has more dispersion among pathologies. This is consistent with the fact that CTV_MR is based on anatomic references and not on pathological tissue. Parotid gland and nasopharynx are the cases with the highest DSC difference comparing CTV_HR and CTV_MR. Also, parotid gland and oropharynx targets show the same behavior: CTV_HR was contoured very differently from each other and very different from the instructor, whereas CTV_MR contours are more similar to each other. The remaining pathologies have both targets and similar performance values among them. Cervical metastases and larynx are the only cases in which SD for CTV_MR is less than for CTV_HR.

Figure 4C shows GTV or CTV_HR DSC dependence on having a supporting diagnostic image. It seems that the support of another set of images has no influence on the highest target contouring. Results are diverse by pathology and/or target which hide the help of these supporting images.

Volume analysis

Table 2 illustrates the statistics summary of data related to volume relative (ΔV) analysis detailed by

Table 1 Median, mean, standard deviation (SD), minimum (Min), maximum (Max), and coefficient of variation (CV) of DSC stratified by pathology and target volumes

Pathology	Targets	N	Median	Mean	SD	Min	Max	CV
Cervicalmet	CTV_MR	13	0.59	0.53	0.13	0.28	0.65	0.25
Cervicalmet	CTV_HR	13	0.54	0.50	0.11	0.30	0.64	0.21
Larynx	CTV_MR	14	0.74	0.72	0.08	0.49	0.80	0.11
Larynx	CTV_HR	14	0.79	0.78	0.05	0.67	0.84	0.06
Larynx	CTV_LR	14	0.72	0.72	0.04	0.62	0.78	0.05
Larynx	GTV	14	0.86	0.84	0.04	0.73	0.89	0.05
Nasopharynx	CTV_MR	10	0.67	0.64	0.08	0.54	0.70	0.12
Nasopharynx	CTV_HR	10	0.47	0.48	0.07	0.38	0.64	0.15
Nasopharynx	GTV	10	0.65	0.61	0.14	0.28	0.79	0.23
Oral	CTV_MR	13	0.72	0.72	0.06	0.58	0.80	0.09
Oral	CTV_HR	13	0.55	0.56	0.10	0.43	0.70	0.18
Oropharynx	CTV_MR	20	0.73	0.70	0.08	0.48	0.79	0.11
Oropharynx	CTV_HR	20	0.63	0.60	0.18	0.20	0.80	0.31
Oropharynx	GTV	20	0.62	0.58	0.17	0.07	0.75	0.29
Paranasal	CTV_MR	16	0.69	0.67	0.11	0.43	0.82	0.17
Paranasal	CTV_HR	16	0.70	0.66	0.15	0.39	0.82	0.22
Parotid	CTV_MR	16	0.68	0.66	0.10	0.46	0.78	0.15
Parotid	CTV_HR	16	0.37	0.41	0.25	0.00	0.77	0.62
Parotid	CTV_LR	16	0.66	0.66	0.06	0.52	0.74	0.10
Parotid	GTV	16	0.36	0.38	0.25	0.00	0.73	0.65
SBRT	GTV	16	0.52	0.46	0.20	0.03	0.66	0.45

Abbreviations: CTV_HR = high-risk clinical target volume; CTV_LR = low-risk clinical target volume; CTV_MR = medium-risk clinical target volume; GTV = gross tumor volume; N = number of cases.

targets. Mean ΔV and SD values for GTV, CTV_HR, CTV_MR, and CTV_LR were 0.30 ± 0.39, 0.26 ± 0.75, 0.57 ± 0.44, and 0.01 ± 0.26, respectively. Global ΔV mean and SD obtained were 0.21 ± 0.52. CV mean for all pathologies was 1.33 (−30.91; 54.3).

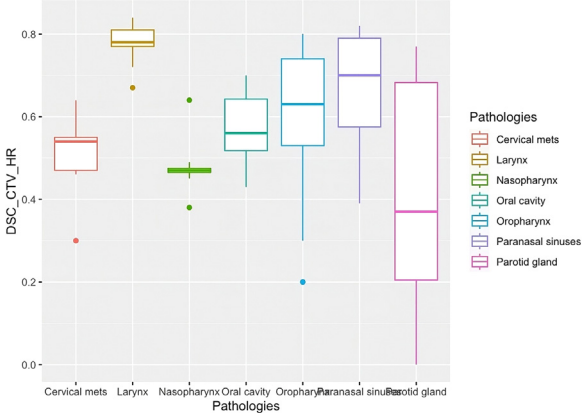

Figure 3 Box plot graph of CTV_HR DSC mean values and their SD divided by pathologies.
Abbreviations: CTV_HR = high-risk clinical target volume; DSC = dice similarity coefficient.

Figure 4D shows ΔV divided by pathology and targets. Parotid gland has a great **ΔV** dispersion among students and a high mean **ΔV** in comparison to the rest of the pathologies. Cervical mets and larynx show higher **ΔV** dispersion in CTV_MR than CTV_HR, performance consistent with DSC analysis. Moreover, as in DSC analysis, larynx contours show the best agreement among pathologies. Oropharynx points show that CTV_HR has remarkably more dispersion than CTV_MR. These results are quite related to variations found in DSC analysis (Fig. 4).

Once again CTV_HR parotid gland has the highest ΔV compared with the rest of pathologies (Fig. 4E). CTV_HR ΔV dispersion, excluding parotid targets, is not higher than CTV_MR ΔV dispersion, which is the main result contradicted by the findings in Fig. 4E.

Regarding the influence of MRI or PET/CT, Fig. 4F shows the ΔV variation with and without the use of supporting images.

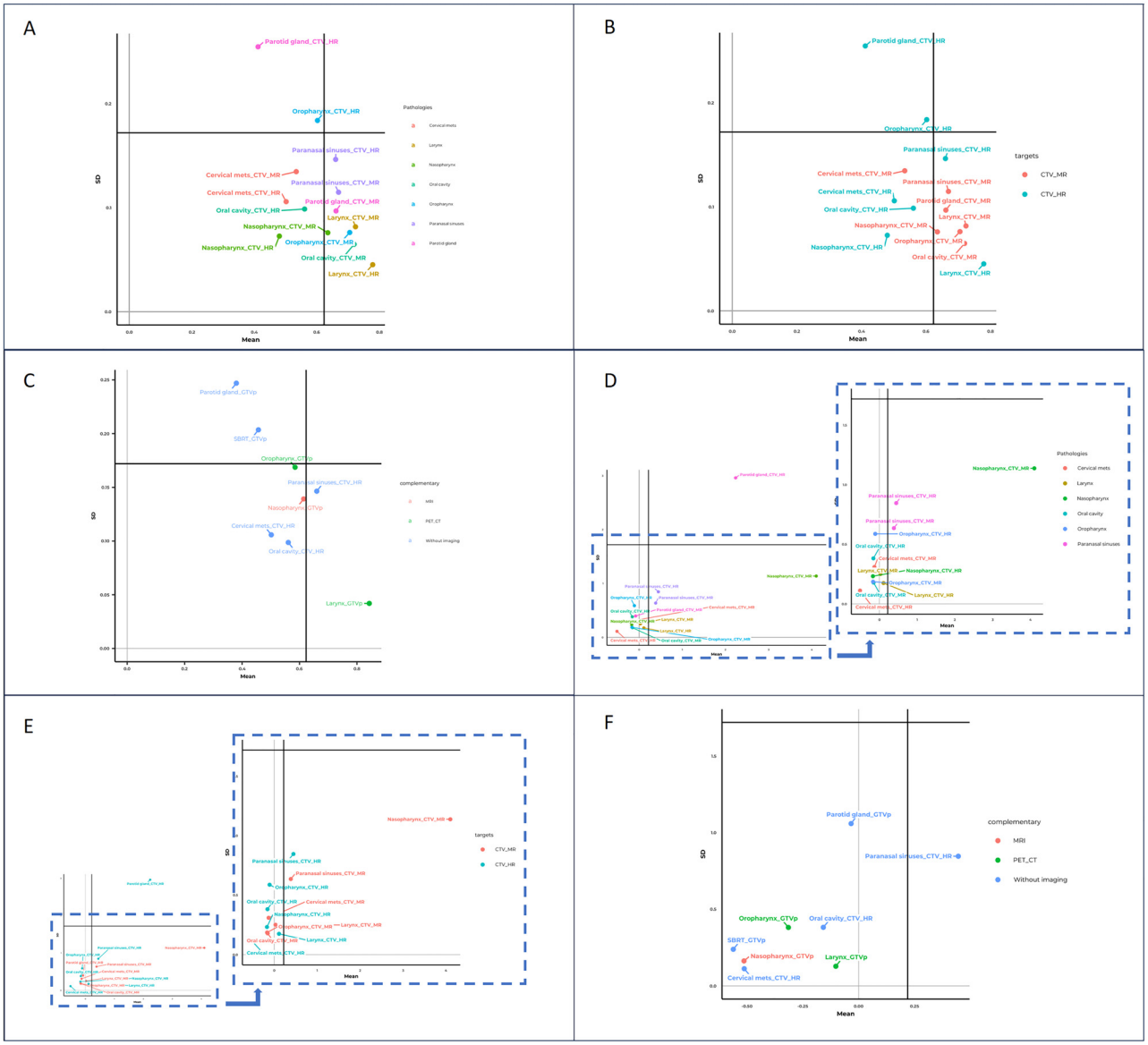

Figure 4 (A) DSC detailed by pathology and target (CTV_HR and CTV_MR). (B) DSC detailed by target (CTV_HR and CTV_MR). (C) DSC detailed by supporting images. (D) ΔV divided by pathology and target (zoom detail excluding parotid gland case). (E) ΔV divided by target (zoom detail excluding parotid gland case). (F) ΔV detailed by supporting images.
Abbreviations: CTV_HR = high-risk clinical target volume; CTV_MR = medium-risk clinical target volume; DSC = dice similarity coefficient.

Discussion

The current project addresses a systematic study of the variability of target contouring in H&N cancer. Other previous comparisons consulted in the bibliography usually refer to a single pathology or to very controlled situations.[11-16]

It is important to emphasize that the analysis does not look for the small differences between operators in contouring details but rather the individual capacities to solve a real situation with the same information. Other authors have studied the influence of teaching using a pre-post comparison[17] or the concordance of resident contours with faculty physician contours.[2]

In this work, we prefer to focus on a more general case. All the students had previous specific training on contouring, oncological experience, and the general clinical information of the cases. Therefore, if the situation were considered a surrogate of a radiation therapy plan of the same cases carried out by different professionals, then, the results of the variability of the contours could be close to a real scenario.

One limitation of the study is that there was a dissimilar previous training of the students enrolled in the course. This can lead to an inconsistency in the delineation of the same case. Moreover, clinical cases were previously presented by the instructors, and students could choose 5 of 8 to get the course certificate, which would

Table 2 Median, mean, standard deviation (SD), minimum (Min), maximum (Max) and coefficient of variation (CV) of ΔV stratified by pathology and target volumes

Pathology	Targets	n	Median	Mean	SD	Min	Max	CV
Cervicalmet	CTV_MR	13	−0.30	−0.13	0.31	−0.49	0.36	−2.35
Cervicalmet	CTV_HR	13	−0.49	−0.51	0.11	−0.69	−0.39	−0.22
Larynx	CTV_MR	14	0.07	0.03	0.25	−0.55	0.32	8.73
Larynx	CTV_HR	14	0.09	0.11	0.18	−0.11	0.52	1.61
Larynx	CTV_LR	14	−0.02	0.00	0.17	−0.34	0.30	54.30
Larynx	GTV	14	−0.11	−0.10	0.13	−0.35	0.19	−1.25
Nasopharynx	CTV_MR	10	3.66	4.10	1.14	2.92	6.29	0.28
Nasopharynx	CTV_HR	10	−0.22	−0.17	0.23	−0.42	0.08	−1.36
Nasopharynx	GTV	10	−0.45	−0.51	0.16	−0.82	−0.27	−0.32
Oral	CTV_MR	13	−0.27	−0.15	0.18	−0.33	0.08	−1.16
Oral	CTV_HR	13	−0.31	−0.16	0.38	−0.58	0.76	−2.41
Oropharynx	CTV_MR	20	−0.22	−0.17	0.19	−0.42	0.17	−1.12
Oropharynx	CTV_HR	20	−0.28	−0.11	0.59	−0.68	1.79	−5.36
Oropharynx	GTV	20	−0.40	−0.32	0.38	−0.78	0.79	−1.21
Paranasal	CTV_MR	16	0.29	0.38	0.64	−0.64	1.84	1.66
Paranasal	CTV_HR	16	0.15	0.45	0.85	−0.64	1.99	1.90
Parotid	CTV_MR	16	−0.14	−0.08	0.40	−0.56	1.04	−4.85
Parotid	CTV_HR	16	1.79	2.24	2.96	−0.52	9.87	1.32
Parotid	CTV_LR	16	0.04	0.03	0.36	−0.52	0.76	11.15
Parotid	GTV	16	−0.43	−0.03	1.06	−0.84	2.47	−30.91
SBRT	GTV	16	−0.55	−0.56	0.24	−0.98	−0.14	−0.4

Abbreviations: CTV_HR = high-risk clinical target volume; CTV_LR = low-risk clinical target volume; CTV_MR = medium-risk clinical target volume; GTV = gross tumor volume; N = number of cases.

result in them choosing the easiest cases to contour or the ones that would have had the clearest explanation of the limits of each target. This would result in cases with lower contouring variability from the outset.

Despite the existence of other metrics (Jaccard Index, Concordance Index, Mean Surface Distance, 95% Hausdorff Distance, etc) for analyzing differences in contouring,[1,6,18] DSC is an especially useful metric for comparison with another study,[2] simple and easy to calculate from TPS and widely used in medical imaging. One of the disadvantages of using this metric is that DSC is linked to structure volume, and smaller structures could yield smaller DSC with an equivalent volume difference.[19]

Global DSC mean value (0.61 ± 0.12) gives us an indication that the contours are not very similar to our reference (instructor). Our results show a situation far from Fig. 2A and that will generate a large uncertainty in the quality of a radiation therapy treatment. Global CV mean value (0.25) was under one, indicating a low SD and a slight variability between the set of contours.

DSC mean value is growing with volume risk, being minimum for GTV and maximum for CTV_LR. This is an expected result because of the anatomic and guide-based characteristics of less risk volumes (CTV_MR and overall CTV_LR) versus pure pathologic ones (CTV_HR and overall GTV).

In general, the volumes of the neck node levels, especially CTV_LR, have less variability than the GTVs. This can be facilitated by the common use among specialists in radiation therapy oncology of different contouring guides agreed on by the main organizations.[3,5]

Due to the reduced variability in CTV_LR (mean DSC and SD 0.69 ± 0.05, mean ΔV, and SD 0.01 ± 0.26) compared with CTV_HR, CTV_MR, and GTV (0.57 ± 0.16 and 0.26 ± 0.75, 0.66 ± 0.09 and 0.57 ± 0.44, and 0.57 ± 0.16 and 0.3 ± 0.39, respectively), a more deep analysis has been discarded.

In relation to Fig. 4B, cervical mets and larynx are the only ones to have a greater dispersion when contouring CTV_MR than CTV_HR. In both cases, a few students included different neck node levels in the CTV_HR and CTV_MR. CTV_HR is mainly based on patient anatomic references, rather than the rest of the cases. Hence, it is reasonable to expect a similar or better coincidence with CTV_MR, based on anatomic references as well.

Variability in structure volumes has obvious implications for tumor coverage or tissue sparing. According to the global obtained SD (0.12), a difference in volume equals to the standard deviation for a CTV_HR volume of 30 cm^3 is 3.6 cm^3. Such a variation, using a typical 3-mm margin, implies a difference in the planning target volume (PTV) subsequent volume of approximately 15% of this target. This could be critical whether it is a defect or an excess of dose.

In our study, the "parotid" location was by far the one that showed the most differences, both between the students and the teacher, and among the students themselves. This may be due to different reasons: it is a postoperative treatment, and the radiotherapy volumes are selected on a tumor bed, which can facilitate larger volumes. Its anatomic peculiarities differentiate it from the rest of the H&N tumors, particularly its relationship with the masticatory space.[20]

One important and reassuring result of this contouring variability study is that DSC and ΔV are related. Both metrics show the same patterns, suggesting that differences could arise from smaller or higher contouring focusing within the same anatomic region rather than from delineating different areas (situation described in Fig. 2C).

The complementary hypothesis 1 has turned out to be false. The presence of a GTV does not contribute to a lower variability in the segmentation of the CTV_HR. This interesting result is, however, difficult to explain with the present data.

An unexpected result is that complementary hypothesis 2 could not be confirmed either: the availability of a supporting PET or MRI is not detected as a factor reducing DSC or volume variability. However, other authors have clearly demonstrated that this type of supporting image does improve the reproducibility and accuracy of contouring.[21-28] This leads one to think that its influence in the cases studied in this work would be less than that of other factors, such as dependency on the person or the difficulty inherent to some pathologies compared with others.

Finally, in this study, we have not investigated the variations in OAR contourings. This variability, reported by van der Veen et al,[19] was significantly lower than that detected in our study of the different targets. In this sense, more effort needs to be made to accomplish further treatment standardization, for example, with artificial intelligence techniques.[29-31]

Conclusions

The results obtained demonstrate the situation that can be found in the contour of patients in the daily routine of a radiation therapy center.

In general, poor DSC agreement has been found between student teachers and great variability between different students.

The differences found show us that there is no uniformity of criteria as would be desirable. This indicates that despite the support of other imaging modalities and the use of guides and recommendations to help contouring, at H&N it is necessary to deepen the training of radiation contouring. Otherwise, serious dosimetric discrepancies could occur in the same case depending on the professional responsible for the treatment.

Disclosures

None.

Acknowledgments

The authors thank the course instructors for their support in compiling the students' contouring and feedback on this project. The authors also thank Navarrabiomed for their collaboration in the translation and publication of this manuscript. Julián Librero was responsible for statistical analysis.

References

1. Trignani M, Argenone A, Di Biase S, et al. Inter-observer variability of clinical target volume delineation in definitive radiotherapy of neck lymph node metastases from unknown primary. A cooperative study of the Italian Association of Radiotherapy and Clinical Oncology (AIRO) Head and Neck Group. *Radiol Med*. 2019;124:682-692.
2. Nissen C, Ying J, Kalantari FK, 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;21:464-472.
3. Jensen K, Friborg J, Hansen CR, et al. The Danish Head and Neck Cancer Group (DAHANCA) 2020 radiotherapy guidelines. *Radioth Oncol*. 2020;151:149-151.
4. Brouwer CL, Steenbakkers RJHM, Bourhis J, et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. *Radiother Oncol*. 2015;117:83-90.
5. Grégoire 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.
6. Grégoire 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.
7. Barghi A, Johnson C, Warner A, Bauman G, Battista J, Rodrigues G. Impact of contouring variability on dose-volume metrics used in treatment plan optimization of prostate IMRT. *Cureus*. 2013;5:e144.
8. Olanloye EE, Ramlaul A, Ntekim AI, Adeyemi SS. FDG-PET/CT and MR imaging for target volume delineation in rectal cancer radiotherapy treatment planning: a systematic review. *J Radiother Pract*. 2022;21:529-539.

9. Bird D, Scarsbrook AF, Sykes J, et al. Multimodality imaging with CT, MR and FDG-PET for radiotherapy target volume delineation in oropharyngeal squamous cell carcinoma. *BMC Cancer*. 2015;15:844.
10. Deantonio L, Beldì D, Gambaro G, et al. FDG-PET/CT imaging for staging and radiotherapy treatment planning of head and neck carcinoma. *Radiat Oncol*. 2008;3:29.
11. Geets X, Daisne J-F, Arcangeli S, et al. Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: Comparison between CT-scan and MRI. *Radiother Oncol*. 2005;77:25-31.
12. Cooper JS, Mukherji SK, Toledano AY, et al. An evaluation of the variability of tumor-shape definition derived by experienced observers from CT images of supraglottic carcinomas (ACRIN protocol 6658). *Int J Radiat Oncol Biol Phys*. 2007;67:972-975.
13. Hong TS, Tomé WA, Harari PM. Heterogeneity in head and neck IMRT target design and clinical practice. *Radiother Oncol*. 2012;103:92-98.
14. Feng M, Demiroz C, Vineberg KA, Eisbruch A, Balter JM. Normal tissue anatomy for oropharyngeal cancer: contouring variability and its impact on optimization. *Int J Radiat Oncol Biol Phys*. 2012;84: e245-e249.
15. Petkar I, McQuaid D, Dunlop A, Tyler J, Hall E, Nutting C. Inter-observer variation in delineating the pharyngeal constrictor muscle as organ at risk in radiotherapy for head and neck cancer. *Front Oncol*. 2021;11: 644767.
16. Gudi S, Ghosh-Laskar S, Agarwal JP, et al. Interobserver variability in the delineation of gross tumour volume and specified organs-at-risk during IMRT for head and neck cancers and the impact of FDG-PET/CT on such variability at the primary site. *J Med Imaging Radiat Sci*. 2017;48:184-192.
17. Lobefalo F, Bignardi M, Reggiori G, et al. Dosimetric impact of inter-observer variability for 3D conformal radiotherapy and volumetric modulated arc therapy: the rectal tumor target definition case. *Radiat Oncol*. 2013;8:176.
18. Wong J, Fong A, McVicar N, et al. Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning. *Radiother Oncol*. 2020;144:152-158.
19. Van der Veen J, Gulyban A, Willems S, Maes F, Nuyts S. Interobserver variability in organ at risk delineation in head and neck cancer. *Radiat Oncol*. 2021;16:120.
20. Loo SW, Martin WMC, Smith P, Cherian S, Roques TW. Interobserver variation in parotid gland delineation: a study of its impact on intensity-modulated radiotherapy solutions with a systematic review of the literature. *Br J Radiol*. 2012;85:1070-1077.
21. Eraj S, Sher DJ. PET/CT: Radiation therapy planning in head and neck cancer. *PET Clin*. 2022;17:297-305.
22. Jensen K, Al-Farra G, Dejanovic D, et al. Imaging for target delineation in head and neck cancer radiotherapy. *Semin Nucl Med*. 2021;51:59-67.
23. Dias Domingues DR, Leech MM. Exploring the impact of metabolic imaging in head and neck cancer treatment. *Head Neck*. 2022;44:2228-2247.
24. Troost EG, Schinagl DA, Bussink J, et al. Innovations in radiotherapy planning of head and neck cancers: role of PET. *J Nucl Med*. 2010;51:66-76.
25. Evesny S, Flaus A, Ailloud A, et al. Therapeutic optimization in head and neck radiotherapy planning: Advocacy for ^{18}FDG PE0T-CT in treatment condition. *Bull Cancer*. 2022;109:1262-1268.
26. Arens AI, Troost EG, Schinagl D, Kaanders JH, Oyen WJ. FDG-PET/CT in radiation treatment planning of head and neck squamous cell carcinoma. *Q J Nucl Med Mol Imaging*. 2011;55:521-528.
27. Minn H, Suilamo S, Seppälä J. Impact of PET/CT on planning of radiotherapy in head and neck cancer. *Q J Nucl Med Mol Imaging*. 2010;54:521-532.
28. Flaus A, Nevesny S, Guy JB, Sotton S, Magné N, Prévot N. Positron emission tomography for radiotherapy planning in head and neck cancer: What impact? *Nucl Med Commun*. 2021;42:234-243.
29. Apolle R, Appold S, Bijl HP, et al. Inter-observer variability in target delineation increases during adaptive treatment of head-and-neck and lung cancer. *Acta Oncol*. 2019;58:1378-1385.
30. Dai X, Lei Y, Wang T, et al. Head-and-neck organs-at-risk auto-delineation using dual pyramid networks for CBCT-guided adaptive radiotherapy. *Phys Med Biol*. 2021;66: 045021.
31. Bilimagga RS, Anchineyan P, Nmugam MS, Thalluri S, Goud PSK. Autodelineation of organ at risk in head and neck cancer radiotherapy using artificial intelligence. *J Cancer Res Ther*. 2022;18:S141-S145.