Radiotherapy and Oncology 137 (2019) 9–15

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Interobserver variability in delineation of target volumes in head and neck cancer

Julie van der Veen [a], Akos Gulyban [b], Sandra Nuyts [a,*]

[a] Department of Oncology, Radiation-Oncology, KU Leuven, University of Leuven, University Hospitals Leuven, Belgium; [b] Department of Radiation Oncology, Europe Hospitals Brussels, Uccle, Belgium

ARTICLE INFO

Article history:
Received 13 July 2018
Received in revised form 25 March 2019
Accepted 4 April 2019
Available online 29 April 2019

Keywords:
Head and neck
Interobserver variability
Contouring
Target volume
Volume delineation
CTV definition

ABSTRACT

Background and purpose: In the last decade precision of radiotherapy treatment execution increased, demanding more accurate delineations to fully exploit these developments. The aim of this study was to identify the extent of interobserver variability in delineation of head and neck cancer (HNC).
Materials and methods: In February 2017 all Belgian radiotherapy departments were invited to complete an online survey and submit clinical target volumes for five HNC reference cases. Clinical target volume of the primary tumour (CTVp) and elective nodal neck (CTVe) were submitted and compared between centres for CTVp and to the CTVe 'gold standard' (CTVeGS). Volume, DICE similarity coefficient (DSC) and median Hausdorff Distance (HD) were measured and calculated.
Results: Fourteen of 22 centres (64%) completed both survey and delineations. They all used delineation guidelines for CTVe and twelve confirmed the use of guidelines of Grégoire et al. Nine centres used CTVp guidelines, although none used the same ones. Median DSC for CTVe comparing centres with CTVeGS ranged between 0.67 and 0.82 and HD50 between 1.7 mm and 2.8 mm. Good agreement was shown for neck level II, III and IV, whilst worst consensus was observed for level Ib, V, VI, VIIa and VIIb. Improvement of DSC and HD50 was observed when the same levels as CTVeGS were selected. Median DSC and HD50 for CTVp ranged between 0.51 and 0.79 and 2.8 mm and 4.1 mm respectively, which both slightly improved when calculating it for only the centres using a 10 mm margin.
Conclusion: Although nearly all participants used identical guidelines for CTVe there were large discrepancies in neck levels selected and volumes delineated. CTVp delineations were also heterogeneous although we expect improvement with implementation of recently published guidelines. Additional teaching in target volume delineation is necessary as this paper demonstrates that availability and implementation of guidelines alone is not enough to guarantee uniform delineation.
© 2019 Elsevier B.V. All rights reserved. Radiotherapy and Oncology 137 (2019) 9–15

Head and neck cancer (HNC) is the seventh most common cancer and cause of cancer-related death worldwide with yearly more than 500 000 new cases and 380 000 deaths [1]. HNC is usually diagnosed in a locally advanced but curable stage for which a multimodal treatment approach is preferred consisting of surgery followed by radio(chemo)therapy or radio(chemo)therapy as definitive treatment. Surgery for locally advanced tumours can be mutilating whereas radio(chemo)therapy holds the potential for better functional outcomes but can cause treatment-related toxicity. Intensification of radiotherapy [2] and/or concomitant chemotherapy [3], have improved survival at the cost of more toxicity [4]. Nevertheless, loco-regional failure rates are high with approximately 30% over 5 years, which is an important cause of morbidity and mortality [5,6]. Great effort is being made to improve the therapeutic ratio by increasing dose to the target volumes (TV) while decreasing dose to organs at risk (OARs). Especially in HNC this can be challenging because of complex anatomy and close proximity between TV and OARs [7]. Since the implementation of more conformal techniques such as intensity modulated radiotherapy (IMRT) and volumetric arc therapy (VMAT), there has been a decrease in toxicity, especially xerostomia [8]. Also intensity modulated proton therapy (IMPT) holds great potential in terms of normal tissue toxicity due to its more favourable dose-depth characteristics, including a sharp localized high dose delivery at the Bragg peak [9]. With these highly conformal techniques correct delineation is even more crucial because inaccuracy here affects all steps downstream and correct delineation is necessary to produce a representative dose plan. Although a lot of effort is being made to deliver the radiotherapy plan as intended with good quality assurance, image guided radiotherapy and immobilisation of the patient, delineation errors remain

* Corresponding author.
E-mail address: sandra.nuyts@uzleuven.be (S. Nuyts).

https://doi.org/10.1016/j.radonc.2019.04.006
0167-8140/© 2019 Elsevier B.V. All rights reserved.

present during the entire treatment which could result in an insufficient radiotherapy dose to the TV or a higher dose to OARs.

International consensus guidelines for delineation of OARs in the head and neck region were published in 2012 [10] after interobserver variability (IOV) had been shown [11]. International consensus guidelines for delineation of the clinical target volume (CTV) of the elective nodal neck (CTVe) have also been published with a last update in 2013, aiming to reduce treatment variations between clinicians [12–15]. The implementation and benefit of these last guidelines however, has not yet been studied. For CTV of the primary tumour (CTVp), no international consensus guidelines were available until recently [16]. A Belgian initiative called PROCAHN (PROject on CAncer of the Head and Neck) was launched with as main objective to identify IOV in delineation of HNC and the reasons for heterogeneity.

Materials and methods

Study design

1. Five HNC cases were randomly chosen to represent different tumour and nodal stages and tumour sites. Post-operative patients were excluded. A detailed description of each case was given with a full description of clinical examination and endoscopy, diagnostic imaging (MRI, CT, PET-CT) and biopsy (Supplementary data 1). Patients underwent a planning CT scan in the supine position with contrast prior to radiation. The CT images were made on a multidetector-row spiral CT scanner (Somatom Sensation Open, 40 slice configuration; Siemens Medical Solutions, Erlangen, Germany). The acquisition parameters were: gantry un-angled, spiral mode, rotation time 1 s, 40 detector rows at 0.6 mm intervals, table speed 21.6 mm/rotation (pitch = 0.9), reconstruction interval 3 mm at Kernel B30s medium smooth and 120 kVp/230 mAs (ref quality mAs) with CareDose4D. The matrix size was 512 × 512, with a pixel spacing of 0.97 × 0.97 × 3.0 mm in the ×, y and z directions, respectively. Gross tumour volume of the primary tumour (GTVp) and pathological lymph nodes (GTVn) were delineated under supervision of a senior radiation oncologist specialized in HNC (SN) to restrict IOV evaluation to CTV delineation.
2. In February 2017, all 25 radiation oncology departments in Belgium were invited to join this project. Participation was not compulsory. Dedicated software (Aquilab Software, Lille, France) was used to set up a secured network between the assessing facility and each participating department. Planning CT with delineated GTVp and GTVn was given. MRI and PET-CT images were not made available as GTV was already delineated and these imaging modalities were not necessary to delineate CTV. CTVp, CTV of pathological lymph nodes, CTVe and OARs had to be delineated by an experienced HNC radiation oncologist.
3. An online survey enquired about the following: which guidelines were used for delineation of CTVp and CTVe and if these needed updating or clarification, which margins were used from GTVp to CTVp and if and how CTVp was modified (Supplementary data 2).
4. The guidelines [12–15] were used to delineate a ‘golden standard’ for CTVe (CTVeGS) depending on the primary tumour site, T and N stage.

Delineation agreement analysis

DICE similarity coefficient (DSC) was calculated because it is a common and intuitive comparison metric. It is calculated by doubling the overlap volume ($V_{overlap}$) and dividing this by the sum of the two volumes (V_1, V_2).

$$\mathrm{DSC} = \frac{2 \times V_{overlap}}{V_1 + V_2} \quad \text{Ideal value} = 1$$

DSC was calculated in a pairwise fashion between all centres for CTVp and between centres and CTVeGS for CTVe, after which a median DSC was calculated per patient. In general, a value >0.6 (>0.8) is considered (very) good [17] although clinical interpretation is difficult as DSC is more forgiving for the same absolute error for larger volumes than smaller volumes. Therefore, additionally to DSC, pairwise Hausdorff Distance (HD) was also determined in order to compare agreement in absolute terms, independent of volume. Generalization of HD is used to measure the difference between two representations of the same object by a certain distance [18]. HD50 and HD95 stand for the minimum distance between contours so that respectively 50% and 95% of the delineated volume is within this distance. For CTVp, HD was used to evaluate agreement of two contours between centres and for CTVe to compare CTVeGS to the delineations from the different centres. For clinical relevance, CTVe volumes were subdivided into the different neck levels for comparison in level selection with CTVeGS. Additionally, distribution of the generalized HD was used to evaluate the geometrical disagreement of the group of pairwise HD comparisons.

Results

Two radiation oncology departments could not take part in this project because of technical reasons, whilst one centre could not upload the delineations. Eventually, fourteen of the 22 remaining centres (64%) responded to the questionnaire and uploaded at least one patient. Eleven centres delineated all five cases leading to a total of 62 submissions.

CTV primary tumour

Survey

Nine institutions confirmed using guidelines for CTVp although these all differed. All institutions declared the use of concentrically isotropic margins from GTVp to CTVp, with an adaptation for anatomical boundaries like bone, air and soft tissue (14, 13 and 5 centres respectively). Thirteen centres responded that CTVp should be modified according to primary tumour site. Thirteen centres concluded that guidelines for delineation of CTVp were necessary (Supplementary data 3).

Margins

Three centres delineated two CTVp volumes, namely a boost (CTVpBoost) and prophylactic (CTVpProph) volume. CTVpBoost was considered for analysis because this would be the volume receiving the same dose as CTVp delineated by centres using only one margin. The margin for CTVpBoost ranged from 5 to 8 mm. The eleven remaining centres used one margin, ranging from 5 to 12 mm of which 10 mm was most common (7 centres) (Table 1). The margins declared in the questionnaire sometimes differed from the margins used, so for our evaluation, the latter were considered.

Anatomical correction

In patient 1 all centres excluded air and bone (mandible, base of skull, hard palate and vertebrae), four excluded the submandibular gland (SMSG) and six excluded surrounding muscles and other soft tissue barriers. One centre expanded CTVp anteriorly towards the base of tongue. All centres excluded air in patient 2, ten excluded bone (vertebra with/without hyoid bone), four excluded the SMSG and nine excluded surrounding muscles and soft tissue. Four expanded CTVp towards the soft palate or part of the mobile tongue. In patient 3, nine of twelve centres excluded air and all

Table 1
Patient characteristics, margin and volume differences in CTVp per centre.

		Patient 1 cT4bN0 Oropharynx Stage IVb	Patient 2 cT2N3b Oropharynx Stage IVb	Patient 3 cT2N0 Supraglottis Stage II	Patient 4 cT3N0 Larynx Stage III	Patient 5 cT2N2b Hypopharynx Stage IVa
Nr. of centres		12	13	12	13	10
GTVp (cm^3)		22.5	20.1	21.6	2.3	14.2
Centre	Margin(mm)			CTVp or CTVpBoost volume (cm^3)		
1	10	82.0	82.8	58.1	36.9	61.5
2*	8 (15)	69.7	–	53.8	12.6	–
3	10	118.3	72.7	94.5	46.4	68.3
4	8	71.8	58.3	63.4	14.3	45.3
5	10	79.2	84.6	145.3	66.1	72.5
6*	6 (10)	67.0	53.3	61.3	51.8	41.4
7	10	113.2	99.6	134.4	87.5	76.1
8	12	116.1	109.2	111.8	24.1	73.1
9	10	99.8	85.8	113.0	61.4	91.6
10*	5 (10)	–	45.9	–	7.2	–
11	5	71.1	57.6	70.5	13.4	30.3
12	5 or 6	–	49.2$^{\alpha}$	–	15.8 $^{\beta}$	–
13	10	107.0	97.0	76.2	–	–
14	10	109.8	104.2	105.7	91.6	84.2
			Median and range			
Volume		90.9	82.8	85.3	36.9	70.4
DSC $^{\text{all centres}}$		0.79 (0.67–0.92)	0.79 (0.59–0.93)	0.71 (0.53–0.90)	0.51 (0.15–0.90)	0.76 (0.49–0.91)
DSC ^{10mm}		0.80 (0.70–0.90)	0.82 (0.63–0.92)	0.72 (0.53–0.88)	0.59 (0.24–0.82)	0.83 (0.75–0.91)
HD50		2.8 (1.1–5.3)	3.1 (1.0–6.6)	3.5 (1.2–6.4)	4.1 (0.7–11.9)	2.8 (1.6–5.5)
HD50 ^{10mm}		2.4 (1.4–3.2)	2.4 (1.2–3.3)	3.2 (2.1–6.3)	3.0 (1.9–4.4)	2.1 (1.6–3.1)
HD95		6.7 (4.1–12.8)	7.1 (3.3–14.0)	9.9 (4.2–32.9)	16.1 (3.0–30.0)	8.5 (4.9–15.2)
HD95 ^{10mm}		6.9 (4.8–11.5)	7.7 (5.3–11.7)	12.7 (4.5–31.2)	10.8 (6.4–18.2)	7.2 (5.2–12.1)

Smallest and largest volumes delineated per patient have been underlined. Volume is given in cubic centimetre (cm^3) and HD in millimetre (mm). Analysis of only the centres using a 10 mm margin shows slight improvement of DSC and HD50. HD95 is also better in patients 4 and 5 but is worse in patients 1–3. This is probably due to modifications to CTVp after expansion.
* two margins were used with in brackets the margin for CTVpProph.
$^{\alpha}$ = 5 mm margin, $^{\beta}$ = 6 mm margin.
TNM staging is given according to the last TNM staging system, TNM8.

excluded the vertebral body although more heterogeneity was seen for the hyoid bone. Seven excluded the SMSGs and eight excluded other soft tissue structures. Six centres expanded CTVp, of which five delineated nearly the entire laryngeal organ and one centre expanded CTVp to also include enlarged cervical nodes, although these were not suspicious on diagnostic imaging. In patient 4, six of thirteen centres delineated the entire laryngeal organ of which only one removed intraluminal air, and two even included the entire hyoid bone. One centre expanded CTVp to include the contralateral side of the larynx too. The other six centres only made an isocentric expansion from GTVp to CTVp of which four excluded air, two excluded the thyroid cartilage, three excluded soft tissue and two made no modifications. All centres excluded bone (vertebra and sometimes hyoid bone) in patient 5, eight excluded air and six excluded surrounding muscles. Six centres extended CTVp further in different directions. Supplementary Fig. 1 firstly shows the delineations for all 5 patients from all centres, secondly only delineations of the centres that used a 10 mm expansion margin and lastly delineations according to the new guidelines [16].

HD and DSC

Median DSC varied between 0.51 and 0.79, HD50 between 2.8 and 4.1 mm and HD95 between 6.7 and 16.1 mm (Table 1). There was generally good agreement when looking at HD50 whilst revealing outliers like patient 4 (small GTVp). Large differences were seen in HD95, with up to 32.9 mm in patient 3. When only considering centres that used a 10 mm margin, median DSC and HD50 improved in all patients, although for HD95 this was not the case. Fig. 1*A* shows the median HD distribution for all patients. Up to HD75, the median HD remains under 5 mm for most patients, followed by a steep increase.

CTV elective neck

Survey

All fourteen departments confirmed using delineation guidelines for CTVe, whilst twelve explicitly referred to Grégoire et al. [12–15]. Seven found that the guidelines would benefit from extra clarification.

Level selection

In patient 1 no centre selected the same combination of levels as CTVeGS and in patients 2 and 5 only three centres did. In patients 3 and 4 this was better, with respectively ten and eleven centres. In general, levels II-IV were selected bilaterally in all patients, by all centres, except once. Regarding the other levels, there was more heterogeneity (Table 2).

Volumes

CTVeGS was always smaller than the median volume delineated by the centres. Patients 2 and 5 had adenopathies, which explains why these had the largest volumes. Supplementary Fig. 2 highlights the differences in neck level delineation for all patients.

HD and DSC

Median DSC ranged between 0.67 and 0.82, HD50 between 1.7 and 2.8 mm and HD95 between 6.1 and 19.2 mm, with a small improvement when the same levels as CTVeGS were selected. Median HD distribution for all 5 patients remained below 5 mm up to HD80 for most patients, followed by a steep increase (Fig. 1B).

Fig. 2 shows identical slices of patient 1 delineated by all institutions, which clearly shows the differences in CTVp and CTVe delineation.

Discussion

With steeper dose gradients it has become possible to shape the delivered dose tightly around the TV, thereby sparing normal tissue and reducing toxicity but with an increased risk of a geographical miss and a recurrence if the TV delineation is not accurate. Although studies have shown that most recurrences in HNC occur in the GTV [5,6,19–21], delineating CTV correctly remains important. Discussions on margin reduction, CTVe dose reduction or CTVp dose escalation can only be started when it has been shown how radiotherapy is actually delivered. Every effort to reduce heterogeneity in delineation should be made as this is the only way to pool multicentre data on outcome and morbidity, improve quality of radiotherapy delivered and achieve a more uniform treatment outcome.

The aim of this study was to investigate to which extend CTV delineation guidelines are used in Belgium and if this results in similar contours. It is clear from the survey that guidelines for CTVp delineation were required because the guidelines mentioned all differed from one another and 93% of centres felt that guidelines were necessary. This is in contrast to CTVe guidelines which were used by all centres and twelve referred to the guidelines from Grégoire et al. [12–15]. Still, 50% found that these consensus guidelines needed clarification.

Table 2
Differences in CTVe neck level selection per patient and differences in volumes delineated.

		Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
Total nr. of centres		11	13	11	13	11
GTVn (cm^3)		–	20.5	–	–	8.1
		Nr. of centres who delineated neck levels				
Ipsilateral	Ia	0	0	0	0	0
	Ib	6	11	0	0	10
	II	11	13	11	13	11
	III	11	13	11	13	11
	IV	11	13	11	13	11
	V	0	10	0	0	11
	VI	0	0	1	2	3
	VIIa	7	8	0	0	11
	VIIb	4	10	0	0	8
Contralateral	Ia	0	0	0	0	0
	Ib	1	3	0	0	0
	II	11	13	11	13	11
	III	11	13	11	13	11
	IV	10	13	11	13	11
	V	0	2	0	0	1
	VI	0	0	1	2	3
	VIIa	6	3	0	0	7
	VIIb	2	1	0	0	3
Nr. of centres = CTVeGS		0	3	10	11	3
Nr. of centres with RR risk		5	7	0	0	9
		Median and range				
Volume ^{CTVeGS}		117.8	168.9	144.2	87.5	225.5
Volume $^{all\ centres}$		141.2 (113.7–215.2)	208.6 (183.5–258.2)	167.0 (134.9–268.9)	108.8 (78.4–199.1)	256.7 (215.7–421.3)
DSC $^{all\ centres}$		0.75 (0.64–0.83)	0.76 (0.62–0.82)	0.82 (0.64–0.89)	0.76 (0.56–0.85)	0.67 (0.59–0.82)
DSC $^{centres=CTVeGS}$		–	0.78 (0.78–0.78)	0.81 (0.64–0.89)	0.78 (0.69–0.85)	0.72 (0.71–0.82)
HD50 $^{all\ centres}$		2.0 (1.5–3.4)	2.0 (1.6–3.4)	1.7 (1.5–3.1)	1.7 (1.2–2.8)	2.8 (2.0–3.7)
HD50 $^{centres=CTVeGS}$		–	1.9 (1.9–2.0)	1.7 (1.5–3.1)	1.7 (1.2–2.6)	2.3 (2.0–2.4)
HD95 $^{all\ centres}$		14.5 (7.7–24.1)	12.2 (6.7–20.2)	6.1 (4.8–26.1)	7.1 (5.8–25.3)	19.2 (9.8–25.7)
HD95 $^{centres=CTVeGS}$		–	7.72 (6.9–7.9)	5.7 (4.8–14.2)	6.8 (6.0–11.6)	18.0 (16.7–18.4)

The difference in number of centres that delineated the neck levels shows there is variability in neck level selection. Using the guidelines from Grégoire et al. [12–15] the neck levels were delineated to create the 'golden standard' (CTVeGS) (underlined in table), to which DSC, HD50 and HD95 were compared. The table shows how many centres delineated the same levels as CTVeGS, also showing how many centres missed neck levels, which could result in a regional recurrence (RR). Volume is given in cubic centimetre (cm^3) and HD in millimetre (mm). When superscript 'centres = CTVeGS' is used, calculations are only done with the centres that delineated the same levels as CTVeGS. Abbreviations: GTVn: GTV nodes.

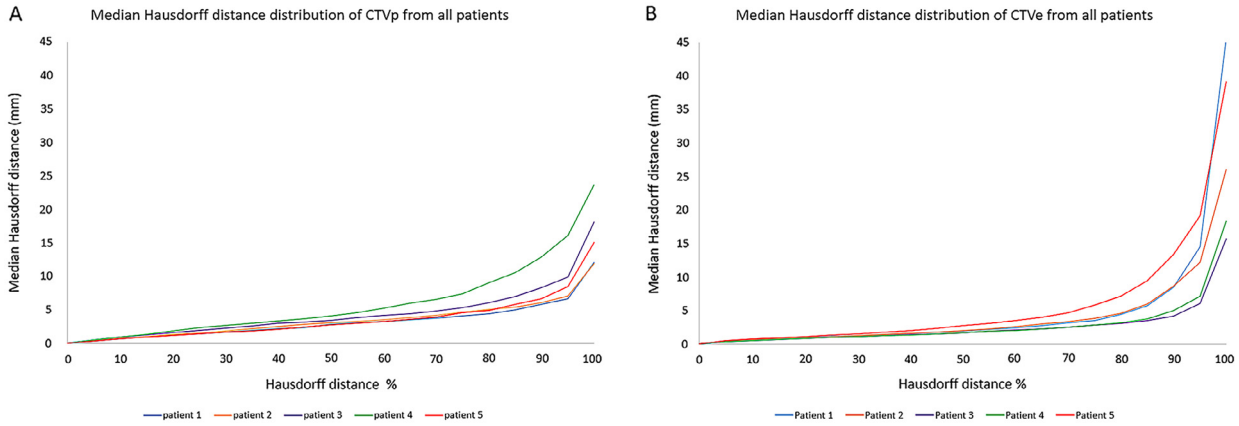

Fig. 1. Median Hausdorff distance for CTVp and CTVe for all patients. (A) The median Hausdorff distance of CTVp between centres is given. There is no significant difference between patients. HD50 is for all 5 patients under 5 mm which means that half of the time, the distance between contours is less than 5 mm. Above HD75 there is an exponential increase in median Hausdorff distance. The maximum difference (HD100) ranges from 11.9 mm (patient 2) to 23.7 mm (patient 4). (B) This shows median Hausdorff distance of CTVe that centres delineated compared to CTVe delineation according to the guidelines of Grégoire et al. (CTVeGS) [12–15]. There is no significant difference between the patients. HD50 is for all 5 patients under 5 mm. Above HD80 there is an exponential increase in median Hausdorff distance with a maximum difference (HD100) ranging from 15.7 mm (patient 3) to 45.6 mm (patient 1).

The lack of CTVp guidelines was reflected in several ways. Firstly, eleven centres used one expansion margin and three used two margins, which would correspondingly receive different radiation doses. Expansion margins also varied (5–12 mm) although 10 mm was most common. The margins the centres declared they used in the questionnaire sometimes differed from the margins actually used in the delineations. This could have occurred by mistake or there could be a difference in theoretical and practical knowledge of delineation. However, a difference in expansion margin alone cannot explain the observed IOV because DSC and HD50 for centres using a 10 mm margin only improved slightly, and HD95 did not improve for all patients. This brings us to the second reason for IOV. Although thirteen centres answered that they adapted CTVp according to tumour site, for example, to compensate for submucosal spread, this was not always performed and also differed strongly (Supplementary Fig. 1B, E, H, K, N). Such inconsistencies were observed in all patients but were particularly present in patient 4 as half of the centres directly delineated the entire laryngeal organ instead of using an expansion margin. Another example is extension to the base of tongue in patient 1 by one centre and in patient 5, six centres extended CTVp beyond geometrical expansion, probably because submucosal expansion is known in hypopharyngeal tumours [22]. Furthermore, correction for anatomical boundaries was done inconsistently concerning the hyoid bone, thyroid cartilage, SMSGs, other soft tissue boundaries and even intraluminal air (e.g. no or in-air safety margin around GTVp). Only vertebrae were consistently excluded. We expect less IOV with the new delineation guidelines as they clearly explain when these structures can be omitted. Overall, not excluding regions where this would be safe, could cause an increase in toxicity, whereas excluding or missing required volumes could result in a geographical miss. In our investigation median HD50s were smaller than 5 mm in all patients with an exponential increase above HD75 for most patients, corresponding well to the clinical observation of differences in modification of CTVp, which leaves room for improvement.

Regarding CTVe, even with (assumedly) same guidelines used, 0, 3, 10, 11 and 3 centres agreed with CTVeGS for patients 1–5 respectively. On the one hand, there is risk of a regional recurrence (RR) if the necessary neck levels are not included. These results were very disappointing with about half of centres failing to delineate the necessary levels in patient 1 and 2 and even nine of eleven centres in patient 5. In patients 3 and 4 all the required levels were selected. On the other hand, if neck levels are delineated while this is not required, this could cause unnecessary toxicity such as additional xerostomia and mucositis for level Ib, skin fibrosis for level V, dysphagia for level VI and xerostomia and dysphagia for level VII. Also, inaccuracy of delineations caused IOV such as exclusion of the SMSGs from level Ib, difference in the posterior border of level V and heterogeneity in delineation of levels VIIa and VIIb. A reason for this varying level selection and delineation could be a misinterpretation of the available guidelines, negligence or conscious adaptation according to the radiation oncologist's clinical experience.

Other studies have also shown variability in HNC CTV delineation. Pettit et al. [23] distributed questionnaires to UK head and neck oncologists in 2010 and showed significant differences in CTVp construction with the use of anatomical and/or geometrical margins. Hansen et al. [24] recently showed more uniform CTVp delineations with the introduction of concentric geometric expansion in the Danish HNC guidelines compared to an anatomical expansion. Their patient (T2N2b oropharynx cancer) was similar to our patient 2 and they reached a median DSC of 0.88, which is better though comparable to the DSC we found (0.82) for centres only using a 10 mm margin. Hansen et al. and our study show that even though the same margin was applied, after modification for primary tumour site and anatomical boundaries, this still resulted in different contours, as is demonstrated with DSC < 1. This finding underlines the importance of clear instructions for modification of CTVp that takes into account patient anatomy, which are offered by the new guidelines [16]. Hong et al. [25] published a study on global IOV in CTV delineation between 20 HNC radiation oncologists. Their reference case (T2N1 oropharyngeal tumour) was comparable to our reference patient 2. Differences with our study are that it was completed before the new CTVe guidelines [15] were available, radiation oncologists used different CTVe guidelines and our study compared CTVe to a 'golden standard'. Similarities between our two studies were identical and good coverage of ipsilateral levels Ib, II and III (85–100%), whereas ipsilateral level V showed heterogeneity in both studies (65 and 77%). Hong et al. had better results for ipsilateral level VIIa (95 vs 62%) whilst our study had better results for contralateral levels II and III (100 vs 60–80%). All other levels showed similar heterogeneity. These results indicate that implementation of guidelines does not suffice to reduce delineation variability. Several ways in which this could be improved are further clarification of existing guidelines, joint

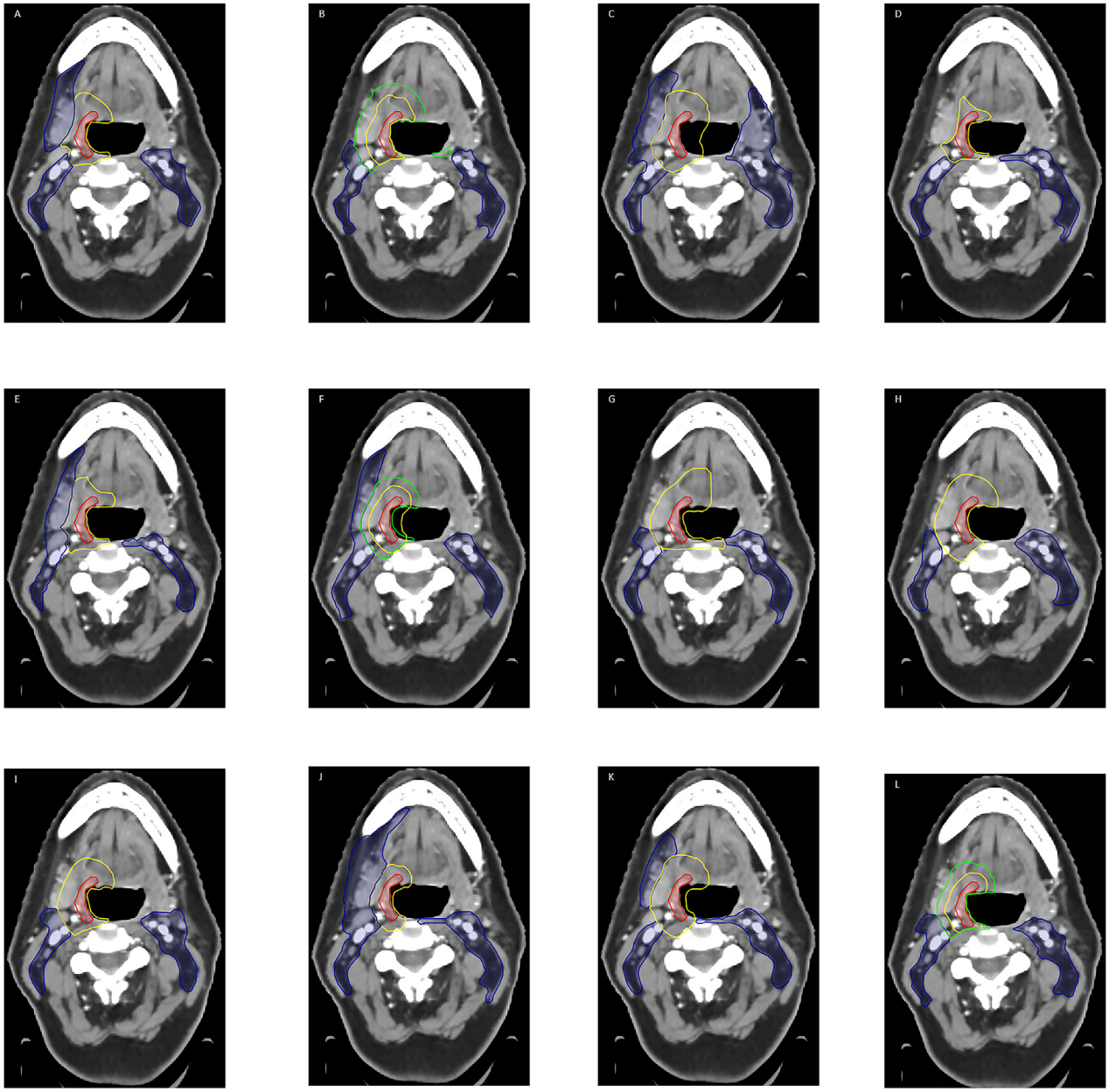

Fig. 2. Target volumes delineated by eleven centres and an example according to the guidelines. This figure shows the differences in CTVp delineation and CTVe level selection and delineation for patient 1, T4bN0 oropharyngeal cancer. GTVp was given (red). CTVp and CTVpBoost are delineated in yellow and CTVpProph in green. CTVe is delineated in blue. Notice the heterogeneity in delineated volumes for this tumour. Two centres delineated two CTV primary volumes. Five centres delineated ipsilateral level Ib, one also delineated contralateral level Ib and five delineated neither. At the bottom right (L), the delineations are shown according to the CTVe guidelines and the newly published CTVp guidelines, from Grégoire et al. [12–16]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

delineation review sessions and systematic training in the form of online educational platforms like FALCON (Fellowship in Anatomic deLineation and CONtouring) which is an educational ESTRO (European SocieTy for Radiation and Oncology) project [26].

Strengths of our study are inclusion of the survey as well as a high response rate (64%) of experienced HNC radiation oncologists. Furthermore, the selected cases reflected daily practice and revealed differences in CTVp and CTVe. CTVeGS was delineated using the consensus guidelines [12–15] which has not been done before and which allowed the levels most at risk of a RR to be highlighted. Alternatively, the effect of the publication of the CTVp guidelines [16] on IOV can be evaluated during a follow-up study because these were not available at the time of this study. GTVp was provided so that pure CTVp variation could be determined as it has been shown that GTVp delineation also shows significant IOV [27,28]. Had GTVp not been given then IOV in CTVp would have been even larger. This however, might also have underestimated IOV together with other factors that could cause therapeutic variability such as differences in planning target volume (PTV) margins, dose prescription and systemic therapy. As participation was not compulsory our results could suffer from response bias. Participating specialists may be used to partaking in trials and follow guidelines strictly, consequently also underestimating IOV [29]. Another limitation may be the use of DSC because it is volume dependent and a clear cut-off value discriminating a good from a poor value is unknown. Therefore HD, an absolute distance based

agreement measure, was added. Although DSC and HD improved when the same margins were used or if the same neck levels were selected, clinical relevance of this improvement still has to be shown. Finally, dosimetric comparisons would have been interesting but could not be performed due to too many unknown variables such as PTV margins and prescribed dose.

In conclusion, even though the first step to reducing delineation heterogeneity is implementation of guidelines, this study highlights that this does not suffice as was demonstrated with incorrect level selection despite implementation of international consensus guidelines. CTVp delineations were also heterogeneous although we expect improvement if consensus guidelines are implemented. The ultimate goal is to improve the quality of radiotherapy delivered, reduce therapeutic variability, take advantage of more precise delivery techniques and pool multicentre data.

Acknowledgements

The authors sincerely appreciate the kind co-operation of The College of Radiotherapy in Belgium for their support, and all radiation oncologists who took part. Without them, this study would not have been possible.

Conflict of interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2019.04.006.

References

[1] Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, Brenner H, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015. JAMA Oncol 2017;3:524–48. https://doi.org/10.1001/jamaoncol.2016.5688.

[2] Bourhis J, Overgaard J, Audry H, Ang KK, Saunders M, Bernier J, et al. Hyperfractionated or accelerated radiotherapy in head and neck cancer: a meta-analysis. Lancet 2006;368:843–54. https://doi.org/10.1016/S0140-6736(06)69121-6.

[3] Pignon J-P, le Maître A, Maillard E, Bourhis J. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): an update on 93 randomised trials and 17,346 patients. Radiother Oncol 2009;92:4–14. https://doi.org/10.1016/j.radonc.2009.04.014.

[4] Nuyts S, Dirix P, Clement PMJ, Vander Poorten V, Delaere P, Schoenaers J, et al. Impact of adding concomitant chemotherapy to hyperfractionated accelerated radiotherapy for advanced head-and-neck squamous cell carcinoma. Int J Radiat Oncol Biol Phys 2009;73:1088–95. https://doi.org/10.1016/j.ijrobp.2008.05.042.

[5] Due AK, Vogelius IR, Aznar MC, Bentzen SM, Berthelsen AK, Korreman SS, et al. Recurrences after intensity modulated radiotherapy for head and neck squamous cell carcinoma more likely to originate from regions with high baseline [18F]-FDG uptake. Radiother Oncol 2014;111:360–5. https://doi.org/10.1016/j.radonc.2014.06.001.

[6] Bayman E, Prestwich RJD, Speight R, Aspin L, Garratt L, Wilson S, et al. Patterns of failure after intensity-modulated radiotherapy in head and neck squamous cell carcinoma using compartmental clinical target volume delineation. Clin Oncol 2014;26:636–42. https://doi.org/10.1016/j.clon.2014.05.001.

[7] Grégoire V, Langendijk JA, Nuyts S. Advances in radiotherapy for head and neck cancer. J Clin Oncol 2015;33:3277–84. https://doi.org/10.1200/JCO.2015.61.2994.

[8] Nutting CM, Morden JP, Harrington KJ, Urbano TG, Bhide SA, Clark C, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial. Lancet Oncol 2011;12:127–36. https://doi.org/10.1016/S1470-2045(10)70290-4.

[9] Holliday EB, Frank SJ. Proton radiation therapy for head and neck cancer: a review of the clinical experience to date. Int J Radiat Oncol 2014;89:292–302. https://doi.org/10.1016/J.IJROBP.2014.02.029.

[10] Brouwer CL, Steenbakkers RJHM, Bourhis J, Budach W, Grau C, Grégoire V, 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. https://doi.org/10.1016/j.radonc.2015.07.041.

[11] Brouwer CL, Steenbakkers RJ, van den Heuvel E, Duppen JC, Navran A, Bijl HP, et al. 3D Variation in delineation of head and neck organs at risk. Radiat Oncol 2012;7:32. https://doi.org/10.1186/1748-717X-7-32.

[12] Grégoire V, Coche E, Cosnard G, Hamoir M, Reychler H. Selection and delineation of lymph node target volumes in head and neck conformal radiotherapy. Proposal for standardizing terminology and procedure based on the surgical experience. Radiother Oncol 2000;56:135–50. https://doi.org/10.1016/S0167-8140(00)00202-4.

[13] Grégoire V, Levendag P, Ang KK, Bernier J, Braaksma M, Budach V, et al. CT-based delineation of lymph node levels and related CTVs in the node-negative neck: DAHANCA, EORTC, GORTEC, NCIC, RTOG consensus guidelines. Radiother Oncol 2003;69:227–36. https://doi.org/10.1016/j.radonc.2003.09.011.

[14] Grégoire V, Eisbruch A, Hamoir M, Levendag P. Proposal for the delineation of the nodal CTV in the node-positive and the post-operative neck. Radiother Oncol 2006;79:15–20. https://doi.org/10.1016/j.radonc.2006.03.009.

[15] Grégoire V, Ang K, Budach W, Grau C, Hamoir M, Langendijk JA, 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;110:172–81. https://doi.org/10.1016/j.radonc.2013.10.010.

[16] Grégoire V, Evans M, Le Q-T, Bourhis J, Budach V, Chen A, 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 Oncolog. Radiother Oncol 2018;126:3–24. https://doi.org/10.1016/j.radonc.2017.10.016.

[17] Bland M. An introduction to medical statistics, 2015.

[18] Cignoni P, Rocchini C, Scopigno R. Metro: measuring error on simplified surfaces. Comput Graph Forum 1998;17:167–74. https://doi.org/10.1111/1467-8659.00236.

[19] Leeman JE, Li J, Pei X, Venigalla P, Zumsteg ZS, Katsoulakis E, et al. Patterns of treatment failure and postrecurrence outcomes among patients with locally advanced head and neck squamous cell carcinoma after chemoradiotherapy using modern radiation techniques. JAMA Oncol 2017;10065:1–8. https://doi.org/10.1001/jamaoncol.2017.0973.

[20] Dawson LA, Anzai Y, Marsh L, Martel MK, Paulino A, Ship JA, et al. Patterns of local-regional recurrence following parotid-sparing conformal and segmental intensity-modulated radiotherapy for head and neck cancer. Int J Radiat Oncol Biol Phys 2000;46:1117–26. https://doi.org/10.1016/S0360-3016(99)00550-7.

[21] De Felice F, Thomas C, Barrington S, Pathmanathan A, Lei M, Urbano TG. Analysis of loco-regional failures in head and neck cancer after radical radiation therapy. Oral Oncol 2015;51:1051–5. https://doi.org/10.1016/j.oraloncology.2015.08.004.

[22] Ho C, Ng W, Lam K, Wei W, Yuen A. Submucosal tumor extension in hypopharyngeal cancer. Arch Otolaryngol Head Neck Surg 1997;123:959–65. https://doi.org/10.1001/archotol.1997.01900090073010.

[23] Pettit L, Hartley A, Bowden SJ, Mehanna H, Glaholm J, Cashmore J, et al. Variation in volume definition between UK head and neck oncologists treating oropharyngeal carcinoma. Clin Oncol 2011;23:654–5. https://doi.org/10.1016/j.clon.2011.07.006.

[24] Hansen CR, Johansen J, Samsøe E, Andersen E, Petersen JBB, Jensen K, et al. Consequences of introducing geometric GTV to CTV margin expansion in DAHANCA contouring guidelines for head and neck radiotherapy. Radiother Oncol 2018;126:43–7. https://doi.org/10.1016/j.radonc.2017.09.019.

[25] Hong TS, Tomé WA, Harari PM. Heterogeneity in head and neck IMRT target design and clinical practice. Radiother Oncol 2012;103:92–8. https://doi.org/10.1016/j.radonc.2012.02.010.

[26] Eriksen JG, Salembier C, Rivera S, De Bari B, Berger D, Mantello G, et al. Four years with FALCON - An ESTRO educational project: achievements and perspectives. Radiother Oncol 2014;112:145–9. https://doi.org/10.1016/j.radonc.2014.06.017.

[27] Riegel AC, Berson AM, Destian S, Ng T, Tena LB, Mitnick RJ, et al. Variability of gross tumor volume delineation in head-and-neck cancer using CT and PET/CT fusion. Int J Radiat Oncol Biol Phys 2006;65:726–32. https://doi.org/10.1016/j.ijrobp.2006.01.014.

[28] Anderson CM, Sun W, Buatti JM, Maley JE, Policeni B, Mott SL, et al. Interobserver and intermodality variability in GTV delineation on simulation CT, FDG-PET, and MR Images of Head and Neck Cancer. Jacobs J Radiat Oncol 2014;1:006.

[29] Wuthrick EJ, Zhang Q, Machtay M, Rosenthal DI, Nguyen-Tan PF, Fortin A, et al. Institutional clinical trial accrual volume and survival of patients with head and neck cancer. J Clin Oncol 2015;33:156–64. https://doi.org/10.1200/JCO.2014.56.5218.