Clinical and Translational Radiation Oncology 36 (2022) 121–126

Contents lists available at ScienceDirect

Clinical and Translational Radiation Oncology

journal homepage: www.sciencedirect.com/journal/clinical-and-translational-radiation-oncology

Delineation uncertainties of tumour volumes on MRI of head and neck cancer patients

Ruta Zukauskaite [a,b,c,*], Christopher N. Rumley [b,d], Christian R. Hansen [b,c,e], Michael G. Jameson [b,f,g,h], Yuvnik Trada [b,f], Jørgen Johansen [a], Niels Gyldenkerne [a], Jesper G. Eriksen [i], Farhannah Aly [b,f], Rasmus L. Christensen [c,e], Mark Lee [b,f], Carsten Brink [c,e], Lois Holloway [b,c,f]

[a] Department of Oncology, Odense University Hospital, Kloevervaenget 9, 5000 Odense, Denmark
[b] Ingham Institute for Applied Medical Research, Liverpool & Macarthur Cancer Therapy Centres, SWSLHD Area Cancer Services, Locked Bag 7103, Liverpool BC, NSW 1871, Australia
[c] Department of Clinical Research, University of Southern Denmark, Odense, Denmark
[d] Department of Radiation Oncology, Townsville University Hospital, QLD Australia
[e] Laboratory of Radiation Physics, Department of Oncology, Odense University Hospital, Kloevervaenget 9, 5000 Odense, Denmark
[f] South Western Sydney Clinical Campuses, University of New South Wales, Sydney, Australia
[g] GenesisCare, NSW, Australia
[h] St Vincent's Clinical School, University of New South Wales, Sydney, Australia
[i] Department of Experimental Clinical Oncology, Aarhus University Hospital, Noerrebrogade 44, Bldg. 5, DK-8000 Aarhus, Denmark

ARTICLE INFO

Keywords:
Inter-observer variation
Target delineation
Head and neck cancer
MRI

ABSTRACT

Background: During the last decade, radiotherapy using MR Linac has gone from research to clinical implementation for different cancer locations. For head and neck cancer (HNC), target delineation based only on MR images is not yet standard, and the utilisation of MRI instead of PET/CT in radiotherapy planning is not well established. We aimed to analyse the inter-observer variation (IOV) in delineating GTV (gross tumour volume) on MR images only for patients with HNC.

Material/methods: 32 HNC patients from two independent departments were included. Four clinical oncologists from Denmark and four radiation oncologists from Australia had independently contoured primary tumour GTVs (GTV-T) and nodal GTVs (GTV-N) on T2-weighted MR images obtained at the time of treatment planning. Observers were provided with sets of images, delineation guidelines and patient synopsis. Simultaneous truth and performance level estimation (STAPLE) reference volumes were generated for each structure using all observer contours. The IOV was assessed using the DICE Similarity Coefficient (DSC) and mean absolute surface distance (MASD).

Results: 32 GTV-Ts and 68 GTV-Ns were contoured per observer. The median MASD for GTV-Ts and GTV-Ns across all patients was 0.17 cm (range 0.08–0.39 cm) and 0.07 cm (range 0.04–0.33 cm), respectively. Median DSC relative to a STAPLE volume for GTV-Ts and GTV-Ns across all patients were 0.73 and 0.76, respectively. A significant correlation was seen between median DSCs and median volumes of GTV-Ts (Spearman correlation coefficient 0.76, $p < 0.001$) and of GTV-Ns (Spearman correlation coefficient 0.55, $p < 0.001$).

Conclusion: Contouring GTVs in patients with HNC on MRI showed that the median IOV for GTV-T and GTV-N was below 2 mm, based on observes from two separate radiation departments. However, there are still specific regions in tumours that are difficult to resolve as either malignant tissue or oedema that potentially could be improved by further training in MR-only delineation.

* Corresponding author at: Department of Oncology, Odense University Hospital, Sdr. Boulevard 29, 5000 Odense C, Denmark.
E-mail address: ruta.zukauskaite@rsyd.dk (R. Zukauskaite).

https://doi.org/10.1016/j.ctro.2022.08.005
Received 22 March 2022; Received in revised form 31 July 2022; Accepted 4 August 2022
Available online 6 August 2022

1. Introduction

Radiotherapy is the primary treatment modality for most patients with loco-regionally advanced head and neck carcinomas (HNC) [1–4].

Treatment planning and delivery of radiotherapy (RT) for head and neck carcinoma (HNC) have become technically advanced [1,5,6]. The first step in planning treatment is pivotal: to identify the macroscopic extension of the local primary gross tumour volume (GTV-T) and regional gross tumour volume in lymph nodes (GTV-N) disease. This is traditionally performed on computed tomography (CT) images since historically CT images are used as the planning system reference. The need to optimally identify the primary tumour volume calls for additional functional and morphologic imaging modalities to support and complement the use of CT for optimising the target identification [7–9].

Magnetic resonance imaging (MRI) has become important in radiotherapy due to its superior soft-tissue contrast that enables tumour visualisation as well as organs at risk (OAR) localisation at different sites [10–12]. MRI in the head and neck (HN) region is valuable for most tumour subsites and the separation of tumour from OARs [12,13]. For these reasons, registration of either a dedicated planning MRI scan or a diagnostic MRI scan with planning CT images is often used in clinical practice to aid more accurate tumour delineation.

Technical evolution in radiotherapy delivery has recently enabled the combination of a linear accelerator and MR scanner (MR-Linac), where both planning and treatment can be performed using MRI as a reference [14]. MR-guided treatment is implemented in several centres using an MR-Linac [15–17], however, there are still problems to be solved (e.g. treatment time in fixation masks, definite contouring guidelines, the real benefit of treatment in MR Linac concerning the reduction of RT doses to many organs at risk; human resources) before MR-adapted treatment can be routinely implemented in clinical practice [12,18,19].

One of the acknowledged uncertainties in identifying the GTV is the divergence between contours of different observers, also called inter-observer variance (IOV). The heterogeneity of contouring the same volumes on CT image datasets in the HN region has previously been investigated [20–24], indicating that IOV may influence treatment outcomes negatively [25]. However, the inter-observer variation in GTV target delineation using only MRI is not well known, and only a few have looked at this [19]. Thus, there is a need to improve the understanding of the inter-observer variation of GTV delineations, particularly when patients with HNC are planned for and treated on an MR-Linac, where all imaging is MR-based.

This study aims to assess the interobserver variability of oncologists in identifying the primary tumour (GTV-T) and nodal (GTV-N) GTV of HNC patients on MRI.

2. Materials and methods

2.1. Patient selection

Thirty-two patients with stage II-IV (TNM 7TH edition) HNC in the oral cavity, oropharynx, hypopharynx, or larynx were prospectively enrolled between 2016 and 2017. The patients were enrolled in one of two studies: ″Outcome prediction of radiotherapy based on MR biomarkers″ at Odense University Hospital, Denmark [26], and ″Evaluation of the Role Of Magnetic Resonance Imaging (MRI) in Mucosal Primary Head and Neck Cancer″ at Liverpool & Macarthur Cancer Therapy Centres, Sydney, Australia [27]. The cohort consisted of 17 Danish patients and 15 Australian patients. All patients had histologically-proven local or loco-regionally advanced HNC (Table 1); they were planned to receive IMRT with or without supplementary systemic treatment using national guidelines [28], or if indicated by individual treatment decisions. Patients treated in Denmark predominantly received mainly moderately accelerated radiotherapy (6 weekly fractions) supplemented with radiosensitiser, whereas Australian patients were treated with five fractions per week (Table 1).

Table 1
Patients' characteristics in two treatment centres.

	Patients from Denmark (n=17)	Patients from Australia (n=15)
Sex:		
Male	13	14
Female	4	1
Primary tumour:		
Larynx	4	4
Pharynx	12	11
Oral cavity	1	0
T-classification (TNM 7th edition):		
1-2	14	6
3-4	3	9
N status:		
N0	2	4
N+	15	11
Treatment dose (Gy)/fractions (fx):		
66-68 Gy/33-34fx	17	14
70Gy/35fx	0	1
Fractionation (fx/week):		
Conventional (5/week)	1	11
Altered (6 or 10/week)	16	4
Systemic treatment :		
None	10	4
Chemotherapy	7	11

In Denmark, MRIs were acquired on a 1.5 T Philips Ingenia (Philips Medical Systems BV, Best, The Netherlands). Patients were positioned in the treatment position without a fixation mask to accommodate the use of the dedicated head and neck coil [26]. Both the T1 weighted (w) and T2w MR images were acquired with voxel size 0.9 × 0.9 × 3 mm^3.

In Australia, patients were scanned on a 3 T Siemens Skyra (Siemens, Erlangen, Germany). Scans were acquired without a fixation mask, using a 20-channel head and neck coil. Axial T2w and T1w Dixon water-only scans were used in the present study, with voxel size 0.9 × 0.9 × 3 mm^3 and 0.98 × 0.98 × 3 mm^3, respectively.

2.2. Target volume delineation

Four senior clinical oncologists from Denmark and four radiation oncologists (two senior and two junior) from Australia with experience in contouring HN patients contoured primary tumour GTV (GTV-T) and nodal GTV (GTV-N) on MRI datasets for all 32 patients. All contours were completed separately, such that contours from other oncologists were blinded.

For each patient, clinicians were provided with 1) synoptic report including clinical examination findings, TNM stage and diagnostic imaging reports such as PET/CT, MRI and/or ultrasound, and 2) contouring protocol based on the validated guidelines for tumour delineation in HN region on MRI [29,30]. Clinicians independently contoured GTV-T and GTV-N(s). For MRI-based contours, T1w without Gd-contrast and T2w images were available and inherently co-registered. To focus the study on inter-observer variation assessment rather than the question of diagnosis, one clinical oncologist (RZ) reviewed the planning CTs and identified the relevant GTV_Ns for delineation with a point marker.

Contouring was performed on treatment planning systems available within the observer's institution. Three observers used Pinnacle v14.0 (Philips, Fitchburg, WI, USA), and five observers used MIM (MIM Software Inc., Cleveland, OH).

Clinician contours and the data sets were exported as DICOM-RT files to MIM, where the contour variation analysis was performed.

2.3. Contouring agreement analysis and statistics

To assess the IOV, a simultaneous truth and performance level

estimation (STAPLE) volume was determined for each volume using MIM. STAPLE is an algorithm, which takes a collection of delineations on an image, and computes a probabilistic estimate of the "true" contour based on expectation–maximisation [31].

The STAPLE volume was computed from all eight observers' volumes. Inter-observer contouring variations were assessed using two contouring variation metrics, DICE Similarity Coefficient (DSC) and mean absolute surface distance (MASD). An overlap-based DICE Similarity Coefficient is a statistical measure of spatial overlap between two volumes, and it is defined as 2x intersection volume divided by the sum of the two delineated volumes. Dice ranges in value from 0 (no overlap) to 1 (perfect overlap) [32]. MASD (Mean Surfaced Distance) is the average geometrical distance of the per voxel shortest distance between two surfaces, ideally zero millimetres [33]. The STAPLE volume was used as the reference volume for both these metrics for each GTV-T and GTV-N.

Descriptive statistics to define the means and standard deviations (std) for DSC and MASD for all contours were performed in MATLAB. The correlation between GTV-T and GTV-N with the STAPLE volumes was assessed using the Spearman correlation coefficient for both DSC and mean surface distance.

4. Results

The majority of the 32 HNC were in the oropharynx (n = 20), mainly loco-regionally advanced. Patient characteristics are shown in Table 1. Differences between the two centres were observed for *T*-status, use of accelerated treatment, and administration of chemotherapy. An equal number of patients (n = 6) in each centre had oropharyngeal HPV p16 positive tumours. Two patients with oropharyngeal carcinoma underwent diagnostic tonsillectomy, and for one patient with laryngeal carcinoma, a partial tumour excision was performed before radiotherapy.

A total of 32 GTV-Ts and 68 GTV-Ns were contoured by each observer. The number of GTV-Ns ranged between zero and nine per patient.

Across the eight observers, the mean GTV-T volume for all patients was 12.1 cm^3 (range 0.68–44.1), and the mean GTV-Ns volume was 4.8 cm^3 (range 0.09–46.9).

The median MASD for GTV-Ts and GTV-Ns across all patients was 0.17 cm (range 0.08–0.39 cm) and 0.07 cm (range 0.04–0.33 cm), respectively (Fig. 1 A and B). A negative correlation was observed between the median MASD and median volumes of GTV-Ts (Spearman correlation coefficient −0.5, p = 0.01). The correlation between median MASD and median volumes of GTV-Ns was not statistically significant.

Median DSC relative to a STAPLE volume for GTV-Ts and GTV-Ns across all patients was 0.73 (range 0.53–0.86) and 0.76 (range 0.51–0.94), respectively. Examples in Fig. 2 represent the lowest (A) and the highest (B) DSC for GTV-Ts and GTV-Ns, respectively. A significant correlation was seen between median DSCs and median volumes of GTV-Ts (Spearman correlation coefficient 0.76, $p < 0.001$) and of GTV-Ns (Spearman correlation coefficient 0.55, $p < 0.001$). Box plots for DSCs/MASD and volumes of GTVs are presented as Supplementary 1A and 1B.

5. Discussion

The use of MR-guided RT has evolved worldwide in recent years with more centres gaining access to MRI and the use of MR treatment systems for daily cancer treatment, including HNC [16,18,34]. Inter-observer variability in tumour delineation on CT image datasets is apparent despite existing international guidelines [21,23]. In this study, we aimed to assess inter-observer variability in GTV delineations for HNC on MRI by eight oncologists and found a median agreement for GTV-T among observers well within the typical PTV margins (3 mm) for HNC [35] and even better for GTV-N; however, a few large outliers were identified.

Previous studies assessing IOV in GTV delineation for HNC were performed using different imaging modalities and using diverse measurements/metrics, hence a direct comparison to this study should be done with caution [36–38]. However, a group of researchers representing MR-Linac Consortium Head and Neck Tumor Site Group, and the Joint Head and Neck Radiotherapy-MRI Development Cooperative had recently submitted a paper representing a similar aim as we and using similar methodologies [19]. In the present study, we chose to use DSC and MASD. The DSC is by no means a perfect measure, however, it is still the most reported metric in contouring variation studies and therefore useful for comparison purposes. The MASD has become the new standard, as the clinical interpretation is more straightforward [39]. Both metrics are considered relative to the STAPLE volume. However, it should be noted that the STAPLE volume is unlikely to provide the actual pathological ground truth [19,31,40,41], but does provide the most consistent approach for determining a ground truth across datasets and studies.

We found that mean DSC and MASD for GTV-Ns were closer to optimal (mean DSC 0.76; MASD 0.11 cm) than for GTV-Ts (mean DSC 0.72; mean surface distance 0.19 cm), likely due to smaller GTV-N volumes than GTV-Ts. This might also be related to the fact that it is often easier to identify the boundaries of nodal volumes due to their capsule than primary tumour volumes that tend to grow infiltratively. In this study, it is also important to note that a marker was placed towards the centre of the nodes to show which nodes should be delineated. As the

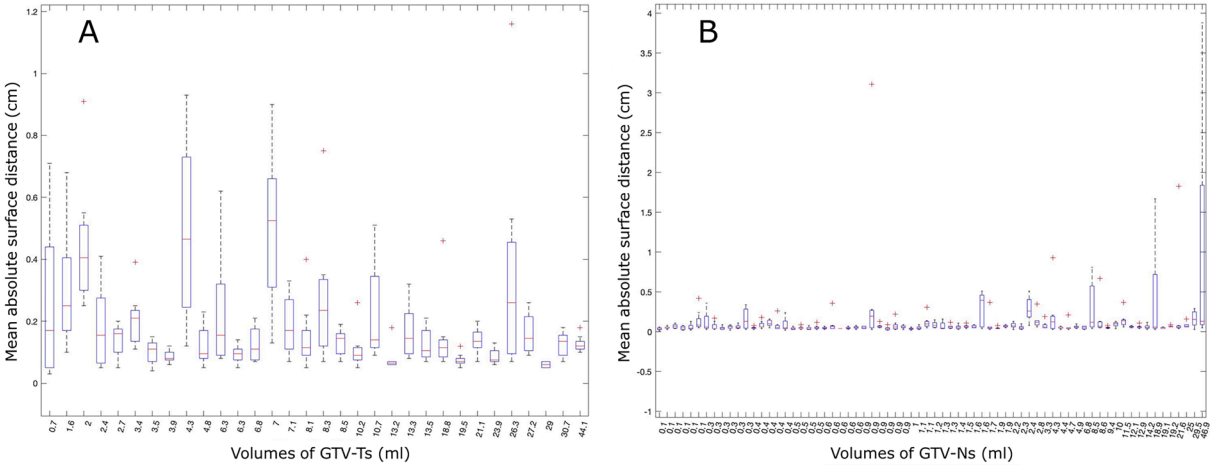

Fig. 1. A: Mean absolute surface distance relative to a STAPLE volume for GTV-Ts across 32 patients among 8 observers; B: Mean absolute surface distance relative to a STAPLE volume for GTV-Ns across 32 patients among 8 observers.

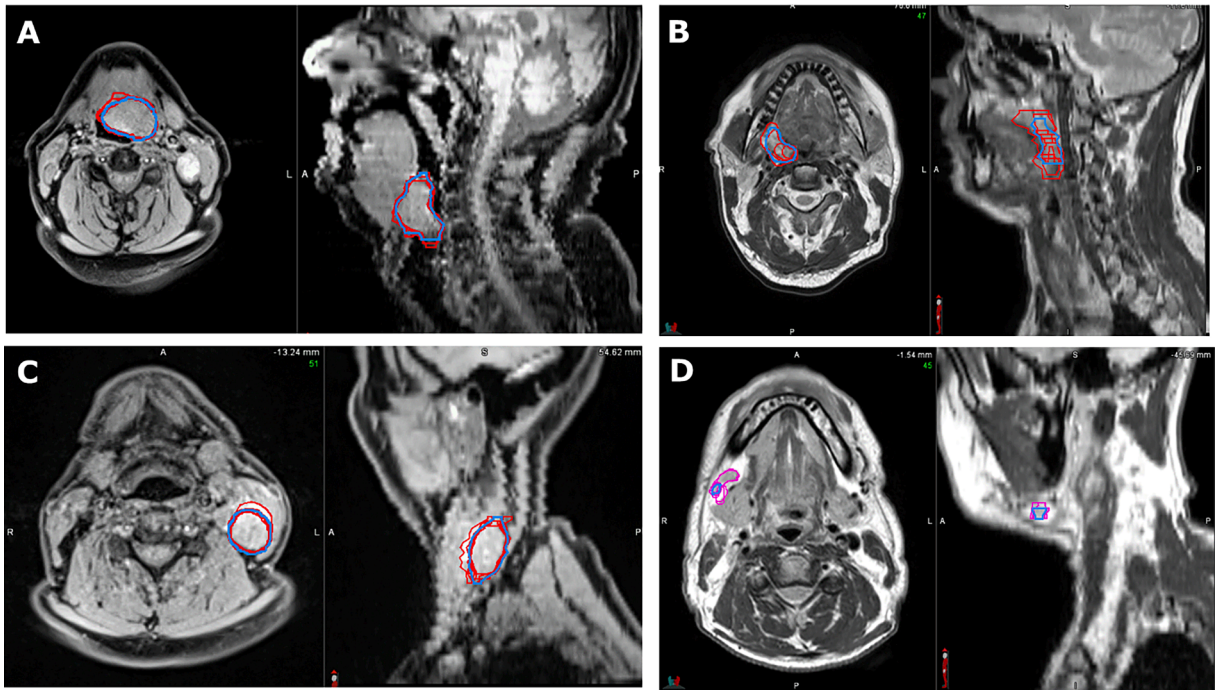

Fig. 2. A: The highest DSC for GTV-Ts, B: The lowest DSC for GTV_Ts, C: The highest DSC for GTV-Ns, D: The lowest DSC for GTV-Ns Contours in red: volumes by 8 observers; contours in blue STAPLE volume *A and C: T1w Dixon water-only MRI; B and D: T2 weighted MRI.* (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

volumes are small, this may also have contributed to greater consistency. For the large GTV-Ns, this capsule boundary became less clear and the contouring variation increased. The uncertainties in GTV-N delineation were mainly related to the patients with three or more lymph nodes that appeared as a conglomerate on some slices; or if one lymph node started on adjacent slices. There were four patients with mean DSC for GTV-T under 0.5. Two of these GTV-Ts were the tonsillar fossa after tonsillectomy, which is often hard to visualise, especially after surgery, and interpretation can vary. Another patient, shown as number 1 in Supplementary 1A, may be considered a real outlier. The patient had a residual vocal cord tumour after the surgery, but the contour of one observer was very different from the STAPLE contour, not even on the same MR slices as other volumes, hence no overlap at all. The study design was aimed to avoid this; however, even with the tumour site specified, post-operation tumour delineation can be difficult.

Previous work looking at inter-observer variation for HNC has focused mainly on comparisons between modalities, particularly between CT and MRI. Cardoso et al. analysed IOV between four observers delineating ten GTVs of head and neck cancer patients on CT/PET, CT/PET/T2W and CT/PET/T2W/DWI image sets. The highest DSC was observed for GTVs delineated on CT/PET (DSC 0.73), whereas the GTV DSC amongst observers for CT/PET/T2W was 0.71 and CT/PET/T2W/DWI – 0.69 [40], which is similar to the results of our study. The volumes using the MRI sequences rather than CT/PET alone were thought to show greater inter-observer variation because of indistinct boundaries at the edges of the delineation volumes. In 2005, Geets et al. published an analysis of IOV of GTV delineations on CT and MRI. Five observers, both radiation oncologists and radiologists, delineated 20 pharyngeal/laryngeal GTVs. Similar to the present study, observers were introduced to the delineation guidelines and used similar MRI sequences. No significant difference in tumour volumes was observed between five observers contouring the same GTVs on MRI [42]. Ten years later, Jager et al. analysed the IOV between three observers delineating laryngeal carcinomas using CT and MRI with gadolinium. Here, the GTV volumes were larger for CT-MRI than for CT alone. Furthermore, adding MR-images to CT showed a decrease in interobserver agreement (using conformity index) compared to the interobserver agreement of the CT-only delineation session. The authors concluded that increased inter-observer agreement and accurate GTV visualisations can only be achieved when clear guidelines for interpretation and delineation on MRI are present [24]. Such guidelines can only be present from the studies where histopathology is used as a gold standard in comparison to delineated GTVs [8,22,29,43]. The most recent work by Cardenas et al. points out a similar need in MR-based delineation guidelines and training in contouring on MRI only. 26 clinicians from 7 countries contoured both GTVs, CTVs and different organs at risk for 4 patients with oropharyngeal cancer as a part of the MRLinac Consortium Head and Neck Tumor Site Group prospective technical benchmarking evaluation (R-IDEAL Stage 0) of human segmentation performance. Similar metrics were used as in the present study but were calculated using both pair-wise and STAPLE approaches. Using the STAPLE algorithm, the MSDs for four GTV-Ts were 1.1, 2.2, 2.2, 2.1 mm, and for two GTV-Ns they were 6.2 and 0.8 mm compared to median MSAD for GTV-Ts and GTV-Ns of 1.7 and 0.7 mm, respectively, in our study. The authors concluded that there is substantial variability between observers' delineation [19] and, similarly to clinicians in our study, argued the need for supplementary information from the other imaging modalities such as PET/CT. However, the observers agreed on invaluable information gained from the description of clinical examinations that is highlighted by other authors [44].

The inclusion of patients in this study was based on two separate projects from Australia and Denmark [26,27]. The patients presented with cancers in different head and neck regions. The majority of patients had pharyngeal malignancies, where MRI traditionally is more used than in the delineation process for laryngeal cancers. The different institutional traditions and possibilities resulted in differences between cohorts; besides the primary aim of projects from where patients were chosen was radiomics features and their analyses. Hence, the geometry was not of the same importance as that would be in the primary delineation study. The inclusion criteria for both studies did not require intravenous contrast or a fixation mask, and only two MRI sequences,

T1w and T2w, were used. These three factors may have an undesirable influence on the results. Performing MR with intravenous contrast would potentially improve the identification of cancerous tissue in the head and neck region and is preferred and recommended to identify tumour volume [12,13]. Some internal motion related to swallowing and respiration in the head and neck region is unavoidable. However, without an immobilisation mask, further patient movement is likely to influence the sharpness of images [12,29]. On the other hand, the lack of immobilisation can allow for dedicated Head and neck MR coils, which improves the image quality. Concerning the MRI sequences, T1w and T2w were used, to enhance tumour and normal tissue contrast [12]. In terms of the history of why sequences were chosen, they were found to be acceptable at the time, and the contrast was avoided where possible, given that the plan was to do multiple scans throughout treatment; it was patient recruitment for the serial scans, also an argument for more consistency for the radiomics where the aim was to duplicate scans. However, with only these two MRI sequences, it can still be difficult to distinguish between the heterogeneous signal intensity of the tumour versus the surrounding tissue as well as a peritumoral inflammation potentially mimicking neoplastic invasion in tumours [8,22]. Diffusion-weighted MRI, which can express high signal intensity with cancerous tissue, is not considered in this study, and could be considered to supplement the target with T1w and T2w sequences [12,40]. However, the DWI has known problems with geometric accuracy and should be used to identify the malignancy but not be used to delineation.

Implementation of MR guided treatment is still in its early stage despite increasing knowledge for some specific anatomical regions, where there are now clear indications [45,46]. The indications to treat cancer patients using MRLinac often are directed to the locations where the soft tissue contrast by the real-time MR guidance provides better opportunities for safe high precision RT with optimal sparing of healthy tissues. An example of prostate cancer is very appropriate: compared to CBCT, onboard MRI can potentially reduce the daily uncertainties in identifying the interface between the posterior part of the gland and the anterior rectal wall or between the prostate apex and penile bulb. But since all the stages in treatment using MRLinac result in a quite prolonged time period, lasting approx.30 min or more, patient inclusion is still a considerable issue and not all patients can be treated despite good indications.GTV changes during a long treatment period, as well as the possibility to decrease actual delivered dose to organs at risk, may justify the use of MR Linacs for some head and neck cancer patients despite long total treatment times [12,18,46,47]; however, these aspects are outside of the scope of this paper. Moreover, guidelines and additional training are needed to improve the consistency of target delineation for the clinical introduction of adaptive radiotherapy based on MRI [19].

HN cancer recurs mainly locally and/or loco-regionally. From the perspective of this study, understanding contouring variations and considering this with respect to the site of relapse may allow oncologists to determine whether the region of contour variation is the region of relapse. This may advance knowledge as to whether tumour biology is the driving force behind the pattern of failures rather than the variation between physicians [48,49]. Understanding contouring variation for MRI for HNC is also necessary to support understanding in considering the evolving knowledge and use of radiomic features on images that have demonstrated potential for predicting tumour response to treatment both on CT and MR images [12,50].

In conclusion, the IOV was found to be on average below 2 mm MASD for primary and nodal disease in head and neck cancer patients between eight observers using MRI alone. For a few patients, significant IOV was found, mainly related to large tumours with unclear boundaries or post-operative conditions. This suggests a need for targeted guidelines, training and contouring quality assurance before MR-based treatment becomes a standard.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ctro.2022.08.005.

References

[1] Baumann M, Krause M, Overgaard J, et al. Radiation oncology in the era of precision medicine. Nat Rev Cancer 2016 Apr;16(4):234–49.
[2] Lacas B, Bourhis J, Overgaard J, et al. Role of radiotherapy fractionation in head and neck cancers (MARCH): an updated meta-analysis. Lancet Oncol 2017;18(9):1221–37.
[3] Lacas B, Carmel A, Landais C, et al. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): An update on 107 randomized trials and 19,805 patients, on behalf of MACH-NC Group. Radiother Oncol 2021 Mar;156:281–93.
[4] Petit C, Lacas B, Pignon JP, et al. Chemotherapy and radiotherapy in locally advanced head and neck cancer: an individual patient data network meta-analysis. Lancet Oncol 2021 May;22(5):727–36.
[5] Gregoire V, Langendijk JA, Nuyts S. Advances in Radiotherapy for Head and Neck Cancer. J Clin Oncol 2015 Oct 10;33(29):3277–84.
[6] Hansen CR, Bertelsen A, Hazell I, et al. Automatic treatment planning improves the clinical quality of head and neck cancer treatment plans. Clin Transl Radiat Oncol 2016 Dec;1:2–8.
[7] Yang J, Beadle BM, Garden AS, et al. A multimodality segmentation framework for automatic target delineation in head and neck radiotherapy. Med Phys 2015 Sep;42(9):5310–20.
[8] Ligtenberg H, Jager EA, Caldas-Magalhaes J, et al. Modality-specific target definition for laryngeal and hypopharyngeal cancer on FDG-PET, CT and MRI. Radiother Oncol 2017 Apr;123(1):63–70.
[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 Nov;4(15):844.
[10] Dirix P, Haustermans K, Vandecaveye V. The value of magnetic resonance imaging for radiotherapy planning. Semin Radiat Oncol 2014 Jul;24(3):151–9.
[11] Kupelian P, Sonke JJ. Magnetic resonance-guided adaptive radiotherapy: a solution to the future. Semin Radiat Oncol 2014 Jul;24(3):227–32.
[12] MRI for Radiotherapy. Planning, deliver and response assessment. Gary L, Heide Uvd, editors. Switzerland: Springer; 2019.
[13] Jensen K, Al-Farra G, Dejanovic D, et al. Imaging for Target Delineation in Head and Neck Cancer Radiotherapy. Semin Nucl Med 2020.
[14] Bertelsen AS, Schytte T, Møller PK, et al. First clinical experiences with a high field 1.5 T MR linac. Acta Oncol 2019;58(10):1352–7.
[15] Lagendijk JJ, Raaymakers BW, Van den Berg CA, et al. MR guidance in radiotherapy. Phys Med Biol 2014 Nov 7;59(21):R349–69.
[16] Bahig H, Yuan Y, Mohamed ASR, et al. Magnetic Resonance-based Response Assessment and Dose Adaptation in Human Papilloma Virus Positive Tumors of the Oropharynx treated with Radiotherapy (MR-ADAPTOR): An R-IDEAL stage 2a–2b/Bayesian phase II trial. Clin Transl Radiat Oncol 2018 Nov;13:19–23.
[17] Raaymakers BW, Jürgenliemk-Schulz IM, Bol GH, et al. First patients treated with a 1.5 T MRI-Linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment. Phys Med Biol 2017;62(23):L41–50.
[18] McDonald BA, Vedam S, Yang J, et al. Initial Feasibility and Clinical Implementation of Daily MR-guided Adaptive Head and Neck Cancer Radiotherapy on a 1.5T MR-Linac System: Prospective R-IDEAL 2a/2b Systematic Clinical Evaluation of Technical Innovation. Int J Radiat Oncol Biol Phys. 2020 Dec 16.
[19] Cardenas CE, Blinde SE, Ng SP, et al. Comprehensive Quantitative Evaluation of Inter-observer Delineation Performance of MR-guided Delineation of Oropharyngeal Gross Tumor Volumes and High-risk Clinical Target Therapy: An R-IDEAL Stage 0 Prospective Study. medRxiv. 2022:2022.01.24.22269596.
[20] Hansen CR, Johansen J, Samsoe E, et al. Consequences of introducing geometric GTV to CTV margin expansion in DAHANCA contouring guidelines for head and neck radiotherapy. Radiother Oncol 2018 Jan;126(1):43–7.
[21] Hong TS, Tome WA, Harari PM. Heterogeneity in head and neck IMRT target design and clinical practice. Radiother Oncol 2012 Apr;103(1):92–8.
[22] Caldas-Magalhaes J, Kooij N, Ligtenberg H, et al. The accuracy of target delineation in laryngeal and hypopharyngeal cancer. Acta Oncol 2015;54(8):1181–7.
[23] van der Veen J, Gulyban A, Nuyts S. Interobserver variability in delineation of target volumes in head and neck cancer. Radiother Oncol 2019 Aug;137:9–15.
[24] Jager EA, Kasperts N, Caldas-Magalhaes J, et al. GTV delineation in supraglottic laryngeal carcinoma: interobserver agreement of CT versus CT-MR delineation. Radiat Oncol 2015 Jan;23(10):26.
[25] Chen AM, Chin R, Beron P, et al. Inadequate target volume delineation and local-regional recurrence after intensity-modulated radiotherapy for human papillomavirus-positive oropharynx cancer. Radiother Oncol 2017 Jun;123(3):412–8.

[26] Christiansen RL, Johansen J, Zukauskaite R, et al. Accuracy of automatic structure propagation for daily magnetic resonance image-guided head and neck radiotherapy. Acta Oncol 2021 May;60(5):589–97.
[27] Rumley CN, Lee MT, Holloway L, et al. Multiparametric magnetic resonance imaging in mucosal primary head and neck cancer: a prospective imaging biomarker study. BMC Cancer 2017 Jul 10;17(1):475.
[28] Jensen K, Friborg J, Hansen CR, et al. The Danish Head and Neck Cancer Group (DAHANCA) 2020 radiotherapy guidelines. Radiother Oncol 2020 Oct;151: 149–51.
[29] Jager EA, Ligtenberg H, Caldas-Magalhaes J, et al. Validated guidelines for tumor delineation on magnetic resonance imaging for laryngeal and hypopharyngeal cancer. Acta Oncol 2016 Nov;55(11):1305–12.
[30] Becker M, Zbären P, Casselman JW, et al. Neoplastic Invasion of Laryngeal Cartilage: Reassessment of Criteria for Diagnosis at MR Imaging. Radiology 2008; 249(2):551–9.
[31] Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging 2004 Jul;23(7):903–21.
[32] Dice L. Measures of the amount of ecologic association between species. Ecology 1945;26:297–302.
[33] Pogson EM, Begg J, Jameson MG, et al. A phantom assessment of achievable contouring concordance across multiple treatment planning systems. Radiother Oncol 2015 Dec;117(3):438–41.
[34] Balermpas P, Krayenbuhl J, Chamberlain M, et al. Feasibility and first experience of MR-guided adaptive radiotherapy for head and neck cancer. Radiother Oncol 2020;152:S443.
[35] Hansen CR, Christiansen RL, Nielsen TB, et al. Comparison of three immobilisation systems for radiation therapy in head and neck cancer. Acta Oncol 2014 Mar;53(3): 423–7.
[36] Jameson MG, Holloway LC, Vial PJ, et al. A review of methods of analysis in contouring studies for radiation oncology. J Med Imaging Radiat Oncol 2010 Oct; 54(5):401–10.
[37] Vinod SK, Jameson MG, Min M, et al. Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies. Radiother Oncol 2016 Nov;121(2):169–79.
[38] Fotina I, Lutgendorf-Caucig C, Stock M, et al. Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. Strahlenther Onkol 2012 Feb;188(2):160–7.
[39] Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging 2015 Aug;12(15):29.
[40] Cardoso M, Min M, Jameson M, et al. Evaluating diffusion-weighted magnetic resonance imaging for target volume delineation in head and neck radiotherapy. J Med Imaging Radiat Oncol 2019 Jun;63(3):399–407.
[41] Allozi R, Li XA, White J, et al. Tools for consensus analysis of experts' contours for radiotherapy structure definitions. Radiother Oncol 2010 Dec;97(3):572–8.
[42] Geets X, Daisne JF, 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 Oct;77(1):25–31.
[43] Daisne JF, Sibomana M, Bol A, et al. Evaluation of a multimodality image (CT, MRI and PET) coregistration procedure on phantom and head and neck cancer patients: accuracy, reproducibility and consistency. Radiother Oncol 2003 Dec;69(3): 237–45.
[44] Thiagarajan A, Caria N, Schöder H, et al. Target Volume Delineation in Oropharyngeal Cancer: Impact of PET, MRI, and Physical Examination. Int J Radiat Oncol Biol Phys 2012;83(1):220–7.
[45] Hehakaya C, der Voort V, van Zyp JR, Lagendijk JJW, et al. Problems and Promises of Introducing the Magnetic Resonance Imaging Linear Accelerator Into Routine Care: The Case of Prostate Cancer. Front Oncol 2020;10:1741.
[46] Lim-Reinders S, Keller BM, Al-Ward S, et al. Online Adaptive Radiation Therapy. Int J Radiat Oncol Biol Phys 2017 Nov 15;99(4):994–1003.
[47] Van Heijst T, Mases J, Meyer Viol SL, et al. MRI linac-based head-and-neck radiotherapy: a new concept for optimal neck sparing. Radiother Oncol 2020;152: S381.
[48] Zukauskaite R, Hansen CR, Grau C, et al. Local recurrences after curative IMRT for HNSCC: Effect of different GTV to high-dose CTV margins. Radiother Oncol 2018; 126(1):48–55.
[49] Andreassen CN, Eriksen JG, Jensen K, et al. IMRT - Biomarkers for dose escalation, dose de-escalation and personalized medicine in radiotherapy for head and neck cancer. Oral Oncol 2018 Nov;86:91–9.
[50] Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014 Jun;3(5):4006.