Radiotherapy and Oncology 194 (2024) 110182

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Improved delineation with diffusion weighted imaging for laryngeal and hypopharyngeal tumors validated with pathology

Hilde J.G. Smits [a,*], Cornelis P.J. Raaijmakers [a], Mischa de Ridder [a], Zeno A.R. Gouw [b], Patricia A.H. Doornaert [a], Frank A. Pameijer [c], Joyce E. Lodeweges [a], Lilian N. Ruiter [d], Koen M. Kuijer [a], Tim Schakel [a], Remco de Bree [e], Jan W. Dankbaar [c], Chris H.J. Terhaard [a], Gerben E. Breimer [d], Stefan M. Willems [f], Marielle E.P. Philippens [a]

[a] Department of Radiotherapy, University Medical Center Utrecht, Utrecht, the Netherlands
[b] Department of Radiotherapy, Netherlands Cancer Institute, Amsterdam, Netherlands
[c] Department of Radiology, University Medical Center Utrecht, Utrecht, the Netherlands
[d] Department of Pathology, University Medical Center Utrecht, Utrecht, the Netherlands
[e] Department of Head and Neck Surgical Oncology, University Medical Center Utrecht, Utrecht, the Netherlands
[f] Department of Pathology and Medical Biology, University Medical Center Groningen, Groningen, the Netherlands

A B S T R A C T

Objective: This study aims to determine the added value of a geometrically accurate diffusion-weighted (DW-) MRI sequence on the accuracy of gross tumor volume (GTV) delineations, using pathological tumor delineations as a ground truth.

Methods: Sixteen patients with laryngeal or hypopharyngeal carcinoma were included. After total laryngectomy, the specimen was cut into slices. Photographs of these slices were stacked to create a 3D digital specimen reconstruction, which was registered to the in vivo imaging. The pathological tumor ($tumor_{HE}$) was delineated on the specimen reconstruction.

Six observers delineated all tumors twice: once with only anatomical MR imaging, and once (a few weeks later) when DW sequences were also provided. The majority voting delineation of session one (GTV_{MRI}) and session two ($GTV_{DW\text{-}MRI}$), as well as the clinical target volumes (CTVs), were compared to the $tumor_{HE}$.

Results: The mean $tumor_{HE}$ volume was 11.1 cm^3, compared to a mean GTV_{MRI} volume of 18.5 cm^3 and a mean $GTV_{DW\text{-}MRI}$ volume of 15.7 cm^3. The median sensitivity (tumor coverage) was comparable between sessions: 0.93 (range: 0.61–0.99) for the GTV_{MRI} and 0.91 (range: 0.53–1.00) for the $GTV_{DW\text{-}MRI}$.

The CTV volume also decreased when DWI was available, with a mean CTV_{MR} of 47.1 cm^3 and a mean $CTV_{DW\text{-}MRI}$ of 41.4 cm^3. Complete tumor coverage was achieved in 15 and 14 tumors, respectively.

Conclusion: GTV delineations based on anatomical MR imaging tend to overestimate the tumor volume. The availability of the geometrically accurate DW sequence reduces the GTV overestimation and thereby CTV volumes, while maintaining acceptable tumor coverage.

Modern conformal radiotherapy techniques, such as intensity modulated radiotherapy, make it possible to maximize the radiation dose to the tumor volume while minimizing the dose to the surrounding tissue. As the precision of dose delivery keeps increasing, the accuracy of the gross tumor volume (GTV) delineation becomes more important. It is one of the biggest sources of uncertainty in radiotherapy treatment [1,2].

In head and neck cancer, GTV delineation using traditional imaging modalities like computed tomography (CT), magnetic resonance imaging (MRI), and [18]F-fluorodeoxyglucose (FDG) positron emission tomography (PET) shows a large inter-observer variability [3–6]. Additionally, studies that validate GTV delineations with histopathology all report an overestimation of the tumor volume, while not achieving complete tumor coverage in all cases [7–9]. This necessitates further expansions of the radiotherapy volumes over the already overestimated GTVs.

In clinical practice, the irradiated clinical target volume (CTV) is created by expanding the GTV with a certain margin to make up for microscopic tumor spread that is not visible on in vivo imaging. In the University Medical Center Utrecht, we employ a CTV margin of 6 mm for laryngeal and hypopharyngeal tumors. This margin is based on a study by Ligtenberg et al. (2017) [7], which compared modality-specific GTVs to the pathological tumor delineations. This study found that GTVs based on anatomical T1 and T2-weighted MRI were smaller than GTVs based on CT, leading to less overestimation of the tumor volume. However, MRI delineations needed a 6.1 mm CTV margin to achieve

* Corresponding author at: University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, the Netherlands.
E-mail address: h.j.g.smits-6@umcutrecht.nl (H.J.G. Smits).

https://doi.org/10.1016/j.radonc.2024.110182
Received 18 December 2023; Received in revised form 16 February 2024; Accepted 18 February 2024
Available online 24 February 2024

adequate tumor coverage, compared to a necessary margin of 4.3 for CT-based delineations.

Using functional MR sequences like diffusion-weighted (DW-) MRI might increase the delineation accuracy on MRI. DW imaging (DWI) reflects the mobility of water molecules in tissue. Because tumors have a high cell density, this mobility is restricted in tumor tissue, creating a high contrast between tumors and their surrounding tissue. Conventional DWI however suffers from geometrical distortions [10]. This problem is especially severe around air cavities, of which there are many in the head and neck area. Recently, a turbo spin echo-based DWI sequence, referred to as the DW-SPLICE, has been implemented that eliminates large distortions [11], making DWI suitable for GTV delineation.

This study aims to determine the added value of DW-SPLICE MRI on the inter-observer variability and accuracy of tumor delineations in laryngeal and hypopharyngeal carcinomas using the pathological tumor delineation as a ground truth.

Methods

Patient selection

Patients with laryngeal or hypopharyngeal squamous cell carcinoma who were scheduled for primary total laryngectomy (TLE) between October 2016 and February 2022 were included in this study. This study was approved by the ethics committee of the University Medical Center Utrecht, the Netherlands, and all patients gave informed consent. A case series of three patients included in this study has been published previously [12].

In vivo imaging

All patients underwent preoperative MRI scans on a 3.0 T Ingenia wide bore MR system (Philips Healthcare, Best, The Netherlands), using two flexible surface coils. The scans were made using an immobilizing five-point head-and-shoulder mask. The following sequences were acquired: a T1-weighted scan before and after gadolinium contrast administration, a T2-weighted scan, and a geometrically accurate turbo-spin-echo DW-SPLICE sequence [11] with three b-values: 0, 200, and 800 s/mm^2. Patients also underwent a preoperative CT scan on a Big Bore CT (Philips Healthcare), which was used for image registration. Details regarding in vivo image acquisition can be found in Supplementary Material 1.

Pathology procedure

After the TLE, the laryngeal specimen was collected from the operating room. A complete overview of the pathological workup can be found in Supplementary Material 2. The specimen was fixated in 4 % formaldehyde for at least 36 h. After fixation, a CT scan was made of the specimen. The ex vivo CT scans were made on a Big Bore CT system (Philips Healthcare) with a slice thickness of 1 mm.

After the post-fixation CT scan, the specimen was embedded in a block of agarose, Supplementary Material 2.1. This block was cut in slices of approximately 3 mm (macro slices) on a conventional meat-slicing machine. The thickness of each individual slice was measured with an electronic caliper. All macro slices were photographed and digitally stacked on top of each other to create a 3D reconstruction of the larynx specimen (Fig. 1, Supplementary Material 2.2).

From the macro slices, 4-µm histological sections were obtained for hematoxylin and eosin (H&E) staining. The tumors were delineated on the digital 3D specimen reconstruction (tumor$_{HE}$), using the H&E stained sections as a reference, Supplementary Material 2.3. These delineations were created under supervision of a dedicated head and neck pathologist (G.E.B.) and served as the ground truth. Tumor delineations on H&E stained sections are a highly reliable standard for accurate delineation of microscopic tumor extent [13,14].

Image registration

The registration method used in this study was previously validated [15]. Because the larynx contains rigid cartilage structures, the shape and size of the enclosed soft tissue were largely maintained. Large deformations were only found in tumors that grew away from the cartilage

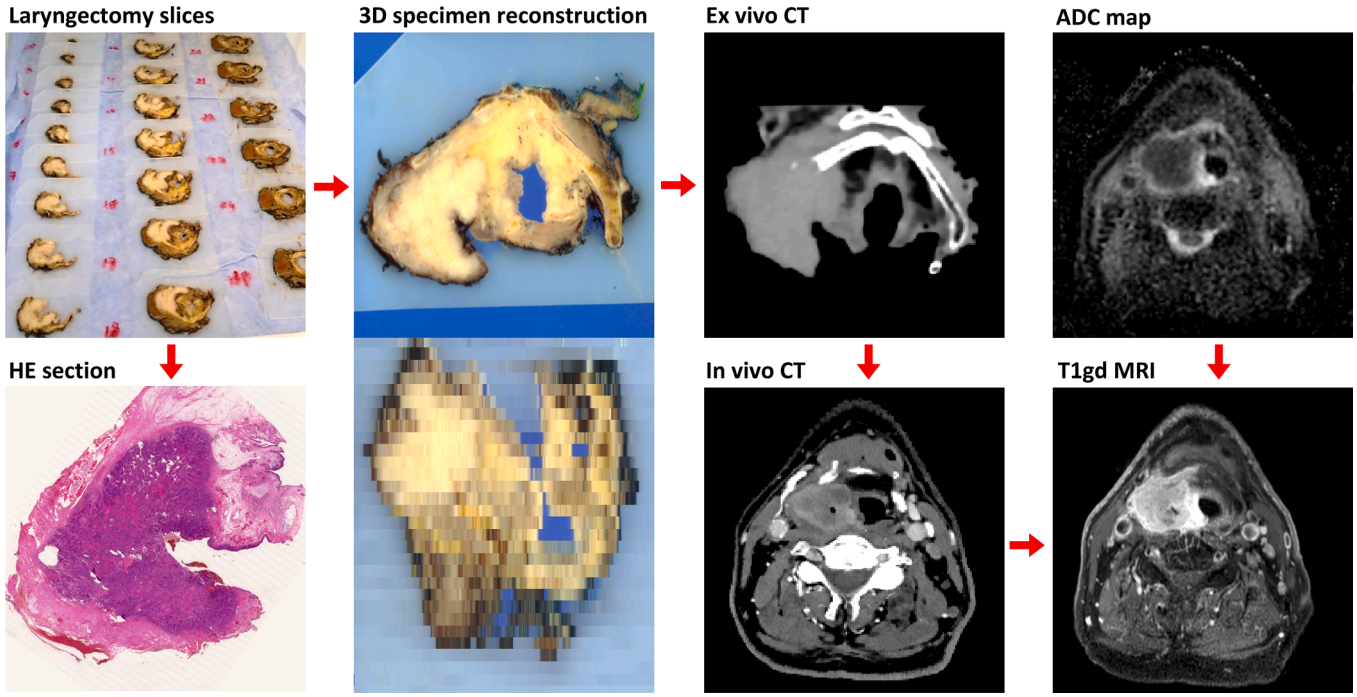

Fig. 1. Digital reconstruction of a laryngectomy specimen and registration process to in vivo imaging. The pathological tumor was delineated on the 3D specimen reconstruction using the hematoxylin and eosin (H&E) stained sections as a reference.

structure [15]. If such deformations were present, the patient was excluded from the analysis.

All initial registrations were done using the Elastix toolbox [16,17]. The reconstructed 3D specimen was automatically registered to the in vivo CT, using the ex vivo CT as an intermediate step, and the in vivo MRI was rigidly registered to the in vivo CT [18] (Figure 1, Supplementary Material 2.4). The pathological tumor delineations were transferred to the MRI and visually inspected to evaluate the automatic registration results. The rigid registrations were manually adjusted if the transferred contour did not match anatomical structures on MRI like cartilage, air cavities, or clear tissue boundaries, Supplementary Material 2.5. In cases where rigid registration did not suffice due to tissue deformations, small adjustments were made to the tumor contour, Supplementary Material 2.6.

In vivo tumor delineations

Five radiation oncologists and one radiologist with various years of experience in delineating head and neck tumors on MRI (range: 6 months to 12 years) participated in the study. Each observer was asked to delineate the GTV of all patients twice. In the first delineation session, only the anatomical MR scans were provided (T1 weighted scans with and without contrast and the T2 weighted scan). In the second delineation session, the DWI b800 images and ADC map were added. There was a minimum of two weeks between the two delineation sessions, and observers were blinded to their previous delineations. Observers received clinical patient information, and endoscopy and radiology reports in order to mimic the clinical routine.

Analysis methods

All delineations were converted to binary masks and processed in MATLAB R2019a (Mathworks, Natick, United States). In order to measure the inter-observer variability, the generalized conformity index (CI_{gen}) was calculated for the delineations made in both delineation sessions. CI_{gen} is defined as the sum of overlapping volumes of each observer pair divided by the sum of the union volumes of each observer pair [19]: $CI_{gen} = \frac{\sum_{pairs ij} |A_i \cap A_j|}{\sum_{pairs ij} |A_i \cup A_j|}$.

For the volumetric and overlap analyses, a majority voting segmentation was created. All voxels that were included in the delineation of at least four observers were included in this segmentation. The majority voting delineations from the first session (GTV_{MRI}) and the second session ($GTV_{DW\text{-}MRI}$) were compared to the tumor delineation based on the H&E stained sections ($tumor_{HE}$).

Four measures were used. The Dice similarity coefficient (DSC) is an overlap measure calculated by dividing the intersect of two delineations by the sum of their volumes: $DSC = 2*\frac{|tumor_{HE} \cap GTV|}{|tumor_{HE}|+|GTV|}$. The sensitivity measures the tumor coverage, the percentage of the tumor volume that is included in the GTV: $sensitivity = \frac{|tumor_{HE} \cap GTV|}{|tumor_{HE}|}$. The positive predictive value (PPV) determines the percentage of the GTV that contains tumor: $PPV = \frac{|tumor_{HE} \cap GTV|}{|GTV|}$. Finally, the 95th-percentile Hausdorff Distance (HD95) was calculated between the $tumor_{HE}$ and the intersection of the $tumor_{HE}$ and GTV. This represents the distance with which the GTV underestimates the $tumor_{HE}$.

To determine the clinical impact of DWI and validate the 6 mm CTV margin used in our clinic, CTVs were created from the majority voting GTVs and compared to the $tumor_{HE}$. The GTVs were isotropically expanded with a 6 mm margin and corrected for anatomical barriers, creating the CTV_{MRI} and $CTV_{DW\text{-}MRI}$.

Results

A total of 20 patients participated in this study, of which 16 were included in the final analysis (Table 1). Two patients were excluded due to large deformations of the specimen between in vivo and ex vivo imaging, Supplementary Material 3. One patient was excluded due to insufficient image quality. Additionally, one patient exhibited spontaneous tumor regression, with only 5 % of vital tumor tissue remaining on histology. Since this caused large discrepancies between the histology and the imaging, this patient was also excluded from analysis. One patient with a T2 tumor was included in this study. This tumor was preoperatively staged as a T4a, which is why this patient received a TLE. The median interval between in vivo MRI and surgery of the remaining patients was 8 days (range: 1 to 23).

Table 1
Patient characteristics.

Patient characteristics		n
Sex	Female	3
	Male	13
Age [years]	Median	68
	Range	52–74
Tumor location	Larynx	9
	Hypopharynx	7
Tumor stage	T2	1
	T3	2
	T4	13

The CI_{gen} from all six observers was comparable for delineation sessions with and without DWI, with a median CI_{gen} of 0.65 (range: 0.41 to 0.74) in the first session and 0.64 (range: 0.41 to 0.72) in the DWI session. The inter-observer variability did not improve when observers could use DWI for their delineations.

When looking at the majority voting volumes, the mean GTV_{MRI} volume was 18.5 cm^3, compared to a mean $GTV_{DW\text{-}MRI}$ volume of 15.7 cm^3 (Fig. 2). This corresponds to a mean volume reduction of 2.9 cm^3 (18.0 %) when DWI was available for the tumor delineations. A complete overview of the results of the GTV analysis can be found in Supplementary Materials 4.

When comparing the majority voting volumes to the pathological tumor volume, $tumor_{HE}$ was much smaller in most cases, with a mean $tumor_{HE}$ volume of 11.1 cm^3 (Fig. 2). The mean volume overestimation of GTV_{MRI} was 7.4 cm^3 (94.9 %) compared to 4.6 cm^3 (53.8 %) for $GTV_{DW\text{-}MRI}$.

The median DSC of the GTV_{MRI} was 0.65 (range: 0.36 to 0.82), compared to 0.72 (range: 0.51 to 0.84) for the $GTV_{DW\text{-}MRI}$. The sensitivity in both sessions was comparable with a median sensitivity of 0.93 (range: 0.61 to 0.99) for GTV_{MRI} and 0.91 (range: 0.53 to 1.00) for $GTV_{DW\text{-}MRI}$, whereas the median PPV increased from 0.50 (range: 0.22 to 0.83) to 0.61 (range: 0.36 to 0.83) for GTV_{MRI} and $GTV_{DW\text{-}MRI}$, respectively. The median HD95 was 0.7 mm for GTV_{MRI} (range: 0 to 7.3 mm) and 1.4 for $GTV_{DW\text{-}MRI}$ (range: 0 to 6.6 mm).

In general, adding DWI had a positive effect on the accuracy of tumor delineations. Areas that may look like tumorous tissue on anatomical MRI sequences were left out of the delineation when they were not visible on the DW-SPLICE, (Fig. 3). However, there was one case in which the $GTV_{DW\text{-}MRI}$ was less accurate than the GTV_{MRI} (Fig. 4). In patient 15, there was cartilage invasion and tumor extension on the anterior side. No diffusion restriction was visible in this part of the tumor on DW-MRI, causing it to be left out of the $GTV_{DW\text{-}MRI}$. Since this area was partly included in the GTV_{MRI}, the sensitivity of the majority voting GTV dropped from 0.83 to 0.53 when DWI was available, meaning that only half of the $tumor_{HE}$ was covered by $GTV_{DW\text{-}MRI}$.

Another notable case is patient 8 (Fig. 5). This patient is the only case in which the pathological tumor volume was underestimated by both the GTV volumes ($tumor_{HE}$: 24.3 cm^3, GTV_{MRI}: 20.7 cm^3, and $GTV_{DW\text{-}MRI}$: 21.5 cm^3). This patient had a relatively long interval of 19 days between MRI and surgery.

Using a 6 mm margin for the CTV delineation, a mean volume reduction of 5.7 cm^3 (13.4 %) was achieved when DWI was available, with a mean CTV_{MR} of 47.1 cm^3 and a mean $CTV_{DW\text{-}MRI}$ of 41.4 cm^3. A

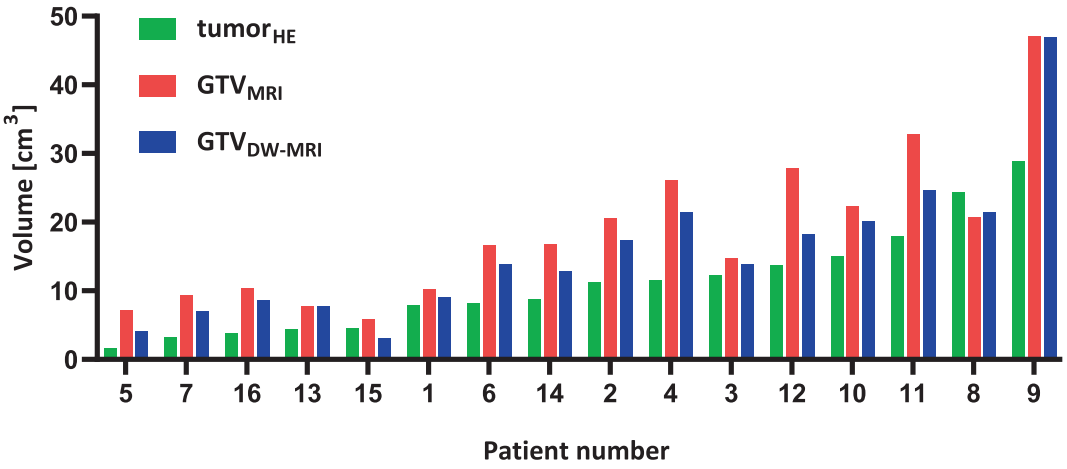

Fig. 2. Comparison of the pathological tumor volume ($tumor_{HE}$) and the majority voting tumor delineation with and without DW-SPLICE (GTV_{MRI} and $GTV_{DW\text{-}MRI}$, respectively). GTV: gross tumor volume, DW-MRI: diffusion-weighted MRI.

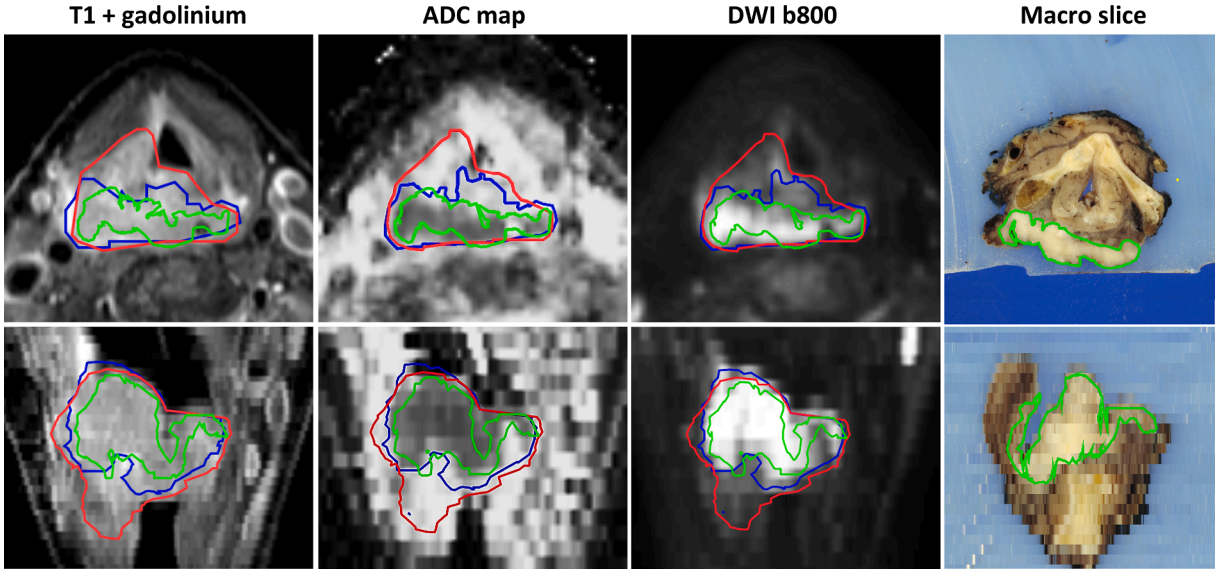

Fig. 3. Tumor volume delineations on MRI, DW-MRI and histopathology of a hypopharyngeal T4 tumor (patient 4) with $tumor_{HE}$ in green, GTV_{MRI} in red and $GTV_{DW\text{-}MRI}$ in blue. The availability of DWI reduced the delineated volume by 4.5 cm^3 (17.6 %) and increased the PPV from 0.43 to 0.52, while a good sensitivity was maintained (0.98 and 0.96, respectively) in this patient. Differences in delineations are caused by differences in slice orientation and slice thickness. ADC: apparent diffusion coefficient, DWI: diffusion-weighted imaging. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

complete overview of the results of the CTV analysis can be found in Supplementary Materials 4. Tumor coverage remained high despite the volume reduction, with complete tumor coverage in 15 and 14 tumors respectively. One patient had incomplete tumor coverage by both the CTV_{MRI} and $CTV_{DW\text{-}MRI}$, although the missed tumor volume decreased when DWI was available (from 0.50 cm^3 to 0.21 cm^3 for CTV_{MRI} and $CTV_{DW\text{-}MRI}$, respectively). In one patient, the tumor was entirely covered by the CTV_{MRI}, but not by the $CTV_{DW\text{-}MRI}$, where a volume of 0.13 cm^3 was missed (patient 15, Fig. 4).

Discussion

This unique study shows that the availability of DW imaging improves tumor delineations. Overestimation of the tumor volume is reduced when observers can use DWI without impacting the sensitivity of the GTV. More importantly, when looking at the CTVs, adding DWI led to an average volume reduction of 5.6 cm^3 while still achieving a high tumor coverage. Such reductions of treatment volume can reduce radiotherapy toxicity, as less healthy tissue is irradiated [20,21]. This is especially relevant if doses can be reduced in the organs at risk. Since head and neck cancers are almost always in the vicinity of multiple organs at risks, like the constrictor muscles, mucosa, and the parotid and mandibular glands [22], reducing treatment volume is an important endeavor. The availability of DWI can therefore aid in reducing radiotherapy toxicities.

Overestimation of the tumor volume on MR has also been found by previous research that compared GTV delineations on MRI to the pathological tumor volume [7,8]. A possible reason for this phenomenon is that edema or inflamed tissue can appear similar to tumor tissue on anatomical MRI sequences, making it difficult to differentiate between them. In this study, we show that the availability of the DW-SPLICE can help exclude these tissues from the GTV and reduce the delineated volume. However, even with DWI, overestimation of tumor volume is still present. This might partially be due to the hesitancy of radiation oncologists who want to prevent underestimation of the tumor volume to ensure good treatment outcomes. Even though we have

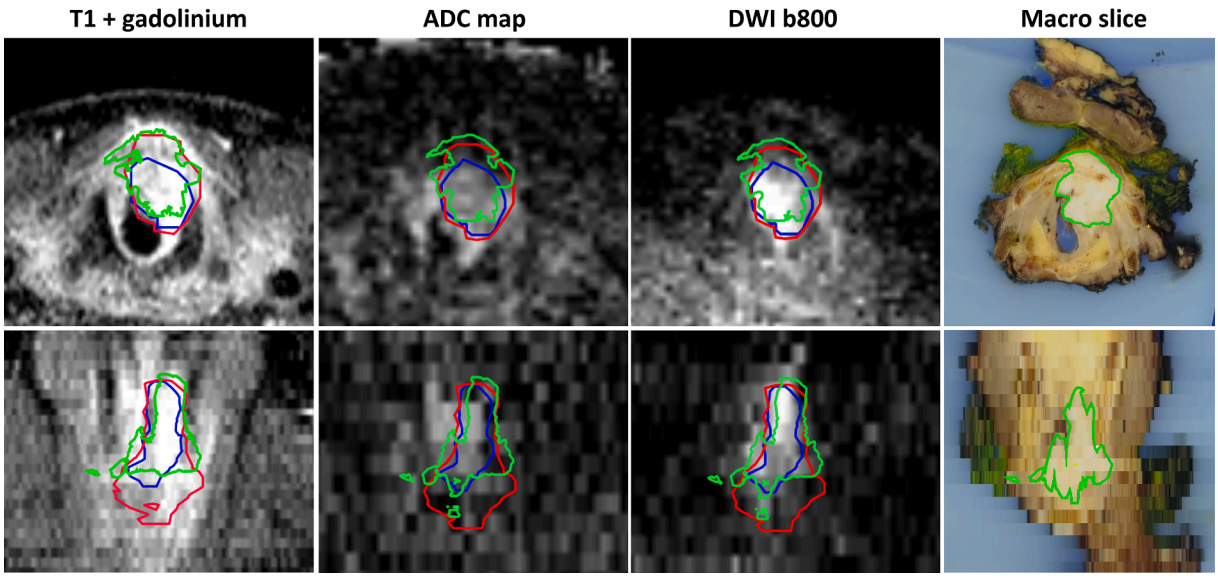

Fig. 4. Tumor delineation on MRI, DW-MRI and histopathology of a glottic T4 tumor (patient 15) with $tumor_{HE}$ in green, GTV_{MRI} in red and $GTV_{DW\text{-}MRI}$ in blue. In this patient, adding DWI resulted in a less accurate tumor delineation. Even though the PPV increased from 0.64 to 0.78 for the GTV_{MRI} and the $GTV_{DW\text{-}MRI}$ respectively, the sensitivity dropped from 0.83 to 0.53. The DWI and ADC images do not adequately depict the tumor extension beyond the cartilage. Differences in delineations are caused by differences in slice orientation and slice thickness. ADC: apparent diffusion coefficient, DWI: diffusion-weighted imaging. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

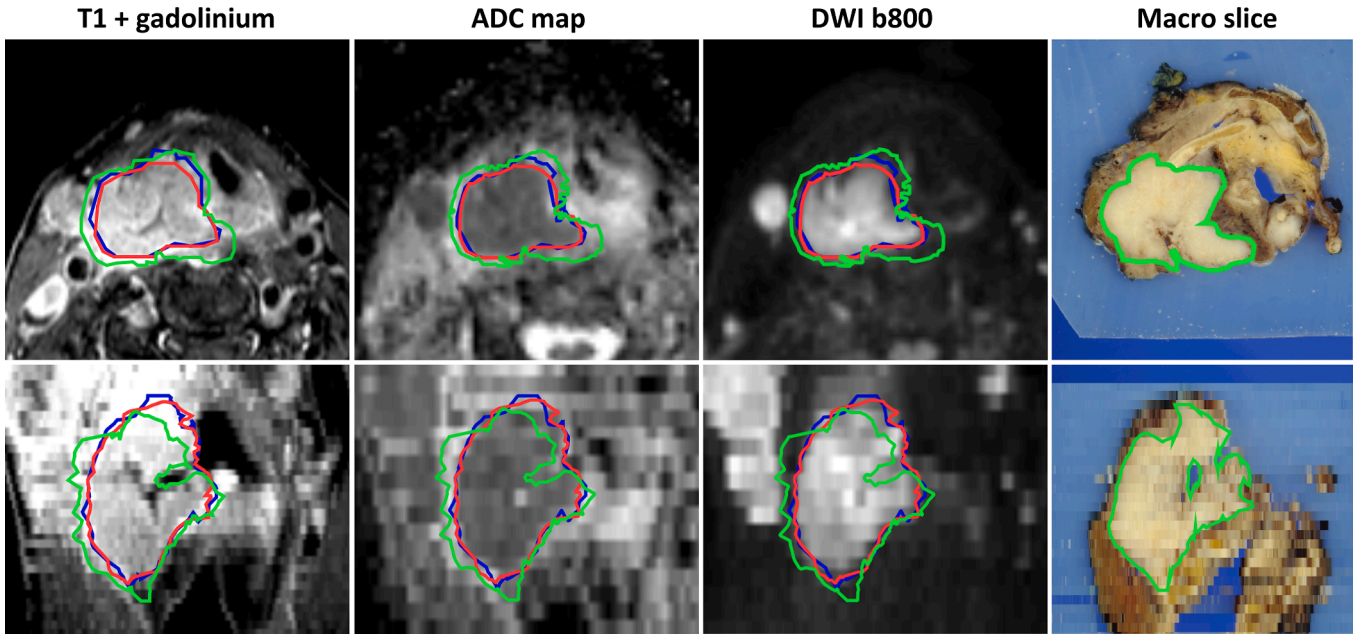

Fig. 5. Tumor delineation on MRI, DW-MRI and histopathology of a hypopharyngeal T4 tumor (patient 8) with $tumor_{HE}$ in green, GTV_{MRI} in red and $GTV_{DW\text{-}MRI}$ in blue. This patient is the only case in which the pathological $tumor_{HE}$ volume exceeded both the GTV volumes. Of note is that this patient had a large lymph node metastasis on the right side of the tumor (left on MRI). Observers were instructed to only focus on the primary tumor, so this metastasis was not included in the delineations. Differences in delineations are caused by differences in slice orientation and slice thickness. ADC: apparent diffusion coefficient, DWI: diffusion-weighted imaging. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

validated delineation guidelines in our clinic [23], we found that these were not always applied, leading to volume overestimation and inter-observer variability. While the overestimation improved with DWI, the inter-observer variability did not. This result was also found in a similar study testing the added value of DW-SPLICE on GTV delineations in rectum carcinoma [24]. The fact that inter-observer variability remains present highlights the need to apply clear delineation guidelines, which have proven to decrease the inter-observer variability of both GTV and CTV delineations [23,25].

While almost all MR delineations overestimated the tumor volume, there was one case in which the pathological tumor volume exceeded both the GTV_{MRI} and the $GTV_{DW\text{-}MRI}$ (patient 8, Fig. 5). This discrepancy is presumably caused by tumor growth in the interval between imaging and surgery. Even though complete tumor coverage was still achieved in this patient by the CTV, treatment delay should be minimized as much as possible. There is a considerable variation between tumor growth rates in head and neck tumors [26,27], and high tumor growth rates have been associated with worse treatment outcomes [26–28].

Additionally, there was one case in which DWI failed to show the complete tumor extent, causing incomplete tumor coverage by the $CTV_{DW\text{-}MRI}$, while CTV_{MRI} did completely cover the tumor (patient 15, Fig. 4). In this patient, the tumor had invaded and extended beyond the thyroid cartilage. While cartilage invasion was present in more cases included in this study, it was usually in a more advanced stage with destruction of the cartilage that was clearly visible on MRI. In patient 15, the cartilage was only invaded in a small area, which was not visible on DWI. Therefore, we recommend not relying on DW imaging when cartilage invasion is present and to make a CT scan to visualize the extent of the cartilage invasion.

The median HD95 doubled when DWI was available, indicating more $tumor_{HE}$ was missed by the $GTV_{DW\text{-}MRI}$ than the GTV_{MRI}. However, the median HD95 distance of the $GTV_{DW\text{-}MRI}$ was 1.4 mm, which is well within the 6 mm CTV margin we apply in our clinic. Moreover, the maximum HD95 decreased when DWI was available and went from 7.3 mm to 6.6 mm for GTV_{MRI} and $GTV_{DW\text{-}MRI}$, respectively.

These maximum HD95 values mean that a 6 mm CTV margin does not lead to complete tumor coverage in all cases. In our dataset, the $CTV_{DW\text{-}MRI}$ completely covers the pathological tumor in 14 out of 16 patients. One of the patient with incomplete coverage was patient 15, where tumor was missed due to cartilage invasion and tumor extension not visible on DWI. Following our recommendations, this patient would have received a CT scan in clinical practice, and the GTV and CTV would have been adjusted accordingly. The remaining patient had incomplete tumor coverage on both CTV_{MRI} and $CTV_{DW\text{-}MRI}$, where 0.50 and 0.21 cm^3 were missed, respectively. This means the tumor coverage increased in this patient when DWI was available.

When determining a suitable CTV margin, we aim for a margin that ensures full tumor coverage in 95 % of patients. A margin that ensures full coverage in all patients would overestimate the treatment volume in most cases and lead to higher toxicity. Incomplete tumor coverage in two out of 16 cases does not meet this criterion, illustrating that while DWI does tend to better visualize the tumor borders, it also has its shortfalls. Besides difficulties in visualizing cartilage invasion, DWI has a low resolution and can be sensitive to motion artifacts. A CT scan can compensate for these shortfalls with its superior resolution and insensitivity to motion. Therefore, we would not recommend omitting CT from the radiotherapy workflow in laryngeal and hypopharyngeal cancer.

A limitation of this study is that patients eligible for radiotherapy treatment typically have lower staged tumors (T1b-T3) than those included in this study (mostly T4 tumors). Small tumors tend to be less well-defined on MRI, so the advantages of DW-MRI reported in this study might not be representative for all patients receiving radiotherapy.

In clinical practice, DWI will likely not be beneficial for the delineation of T1 tumors, as these are often too small for adequate DW imaging. However, three patients with a T2 or T3 tumor were included in our study. In these cases, DWI led to a CTV reduction, while maintaining complete tumor coverage. This could imply that the availability of DWI can be beneficial to the delineation of T2 and T3 tumors as well. However the applicability of our findings to lower-staged tumors should be explored further. In clinical practice, the decision regarding which imaging to use should be taken on a case-by-case basis and the visibility of the tumor on DWI should be considered.

Registration of pathological specimens to in vivo imaging remains challenging. We used a validated registration pipeline [15] and manually adjusted registrations where necessary. However, despite our efforts to optimize the registration, uncertainties will still remain in the final results.

Additionally, we do not account for tumor shrinkage after formalin fixation. Laryngeal resections have the advantage of a rigid cartilage structure, which helps maintain the shape of the enclosed tissue during fixation [15]. In the method we used in this study, the volumetric shrinkage of tissue inside the cartilage was reported to be 3 % [15]. So, while this shrinkage is present, it is not of the same order of magnitude as the volume differences we found between $tumor_{HE}$ and the majority voting GTV delineations (GTV_{MRI} and $GTV_{DW\text{-}MRI}$ overestimated the $tumor_{HE}$ volume by 94.9 % and 53.8 %, respectively).

Conclusions

GTV delineations based on anatomical MR imaging tend to overestimate the tumor volume. The availability of the geometrically accurate DW-SPLICE sequence improves tumor delineations by reducing the GTV and CTV volumes, while maintaining acceptable tumor coverage. This volume reduction might reduce radiotherapy toxicity, as less healthy tissue is irradiated. In a clinical setting, CT should still be included, as cartilage invasion might not always be visible on DWI.

Funding

This research was funded by the Dutch Cancer Society (KWF), project number UU 2011-5216.

CRediT authorship contribution statement

Hilde J.G. Smits: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Cornelis P.J. Raaijmakers:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Mischa de Ridder:** Writing – review & editing, Validation, Conceptualization. **Zeno A.R. Gouw:** Writing – review & editing, Validation, Methodology. **Patricia A.H. Doornaert:** Writing – review & editing, Validation. **Frank A. Pameijer:** Writing – review & editing, Validation, Funding acquisition, Conceptualization. **Joyce E. Lodeweges:** Writing – review & editing, Validation. **Lilian N. Ruiter:** Writing – review & editing, Investigation. **Koen M. Kuijer:** Writing – review & editing, Investigation, Formal analysis, Data curation. **Tim Schakel:** Writing – review & editing, Software, Methodology, Investigation, Data curation, Conceptualization. **Remco de Bree:** Writing – review & editing, Resources. **Jan W. Dankbaar:** Writing – review & editing, Supervision. **Chris H.J. Terhaard:** Writing – review & editing, Validation, Resources, Methodology, Funding acquisition, Conceptualization. **Gerben E. Breimer:** Writing – review & editing, Validation, Supervision. **Stefan M. Willems:** Writing – review & editing, Supervision, Resources, Methodology, Conceptualization. **Marielle E.P. Philippens:** Writing – review & editing, Supervision, Software, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

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.radonc.2024.110182.

References

[1] Njeh CF. Tumor delineation: the weakest link in the search for accuracy in radiotherapy. J Med Phys 2008;33:136–40. https://doi.org/10.4103/0971-6203.44472.
[2] Segedin B, Petric P. Uncertainties in target volume delineation in radiotherapy – are they relevant and what can we do about them? 2016;50(3):254-262. doi: 10.1515/raon-2016-0023.
[3] Jager AA, Kasperts N, Caldas-Magalhaes J, et al. GTV delineation in supraglottic laryngeal carcinoma: interobserver agreement of CT versus CT-MR delineation. Radiat Oncol 2015;10:1–9. https://doi.org/10.1186/s13014-014-0321-4.

[4] 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;48:184–92. https://doi.org/10.1016/j.jmir.2016.11.003.
[5] Van Der VJ, Gulyban A, Nuyts S. Interobserver variability in delineation of target volumes in head and neck cancer. Radiother Oncol 2019;137:9–15. https://doi.org/10.1016/j.radonc.2019.04.006.
[6] Cardenas CE, Blinde SE, Mohamed ASR, et al. Comprehensive quantitative evaluation of variability in magnetic resonance-guided delineation of oropharyngeal gross tumor volumes and high-risk clinical target volumes : an R-IDEAL stage 0 prospective study. Int J Radiat Oncol Biol Phys 2022;113:426–36. https://doi.org/10.1016/j.ijrobp.2022.01.050.
[7] 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;123:63–70. https://doi.org/10.1016/j.radonc.2017.02.005.
[8] Terzidis E, Friborg J, Vogelius IR, et al. Tumor volume definitions in head and neck squamous cell carcinoma – comparing PET/MRI and histopathology. Radiother Oncol 2023;180:109484. https://doi.org/10.1016/j.radonc.2023.109484.
[9] Daisne JF, Duprez T, Weynard B, et al. Tumor volume in pharyngolaryngeal squamous cell carcinoma : comparison at CT, MR imaging, and FDG PET and validation with surgical specimen. Radiology 2004;233:93–100. https://doi.org/10.1148/radiol.2331030660.
[10] Schakel T, Hoogduin JM, Terhaard CHJ, Philippens MEP. Diffusion weighted MRI in head-and-neck cancer: geometrical accuracy. Radiother Oncol 2013;109:394–7. https://doi.org/10.1016/j.radonc.2013.10.004.
[11] Schakel T, Hoogduin JM, Terhaard CHJ, Philippens MEP. Technical note: diffusion-weighted MRI with minimal distortion in head-and-neck radiotherapy using a turbo spin echo acquisition method: diffusion-weighted. Med Phys 2017;44:4188–93. https://doi.org/10.1002/mp.12363.
[12] Ligtenberg H, Schakel T, Dankbaar JW, et al. Target volume delineation using diffusion-weighted imaging for MR-guided radiotherapy: a case series of laryngeal cancer validated by pathology. Cureus 2018;D. https://doi.org/10.7759/cureus.2465.
[13] Jager EA, Willems SM, Schakel T, et al. Interobserver variation among pathologists for delineation of tumor on H & E-sections of laryngeal and hypopharyngeal carcinoma . How good is the gold standard ? Published online 2016. doi:10.3109/0284186X.2015.1049661.
[14] Ligtenberg H, Willems SM, Ruiter LN, et al. Verification of HE-based CTV in laryngeal and hypopharyngeal cancer using pan-cytokeratin. Clin Transl Radiat Oncol 2018;12:21–7. https://doi.org/10.1016/j.ctro.2018.07.003.
[15] Caldas-Magalhaes J, Kasperts N, Kooij N, et al. Validation of imaging with pathology in laryngeal cancer: accuracy of the registration methodology. Int J Radiat Oncol Biol Phys 2012;82. https://doi.org/10.1016/j.ijrobp.2011.05.004.
[16] Klein S, Staring M, Murphy K, Viergever MA, Pluim J. elastix: A toolbox for intensity-based medical image registration. IEEE Trans Med Imaging 2010;29:196–205.
[17] Shamonin DP, Bron EE, Lelieveldt BPF, Smits M, Klein S, Staring M. Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. Front Neuroinf 2014;7:1–15. https://doi.org/10.3389/fninf.2013.00050.
[18] Bol GH, Kotte ANTJ, van der Heide UA, Lagendijk JJW. Simultaneous multi-modality ROI delineation in clinical practice. Comput Methods Programs Biomed 2009;96:133–40. https://doi.org/10.1016/j.cmpb.2009.04.008.
[19] Kouwenhoven E, Giezen M, Struikmans H. Measuring the similarity of target volume delineations independent of the number of observers. Phys Med Biol 2009;54:2863–73. https://doi.org/10.1088/0031-9155/54/9/018.
[20] Al-Mamgani A, Kessels R, Janssen T, et al. The dosimetric and clinical advantages of the GTV-CTV-PTV margins reduction by 6 mm in head and neck squamous cell carcinoma: significant acute and late toxicity reduction. Radiother Oncol 2022;168:16–22. https://doi.org/10.1016/j.radonc.2022.01.013.
[21] Navran A, Heemsbergen W, Janssen T, et al. The impact of margin reduction on outcome and toxicity in head and neck cancer patients treated with image-guided volumetric modulated arc therapy (VMAT). Radiother Oncol 2019;130:25–31. https://doi.org/10.1016/j.radonc.2018.06.032.
[22] 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. https://doi.org/10.1016/j.radonc.2015.07.041.
[23] 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 (Madr) 2016;55:1305–12. https://doi.org/10.1080/0284186X.2016.1219048.
[24] Burbach JPM, Kleijnen JJ, Reerink O, et al. Inter-observer agreement of MRI-based tumor delineation for preoperative radiotherapy boost in locally advanced rectal cancer. Radiother Oncol 2016;118:399–407. https://doi.org/10.1016/j.radonc.2015.10.030.
[25] Bollen H, Gulyban A, Nuyts S. Impact of consensus guidelines on delineation of primary tumor Clinical Target Volume (CTVp) for head and neck cancer : results of a national review project. Radiother Oncol 2023. https://doi.org/10.1016/j.radonc.2023.109915. 109915.
[26] Van Bockel LW, Monninkhof EM, Pameijer FA, Terhaard CHJ. Importance of tumor volume in supraglottic and glottic laryngeal carcinoma. Strahlentherapie und Onkol 2013;189:1009–14. https://doi.org/10.1007/s00066-013-0467-2.
[27] Delahaut G. Prognostic impact of tumor growth velocity in head and neck squamous cell carcinoma treated by radiotherapy : A pilot study. 2019;41:3041-3048. doi:10.1002/hed.25789.
[28] Roldan CS, Chen JJ, Fareed MM, et al. Impact of primary tumor-specific growth rate on treatment failure for nonoropharyngeal head and neck cancers. Laryngoscope 2020;130:2378–84. https://doi.org/10.1002/lary.28393.