

## Original Research Article

# Evaluation of diffusion weighted imaging for tumor delineation in head-and-neck radiotherapy by comparison with automatically segmented $^{18}\text{F}$ -fluorodeoxyglucose positron emission tomography



Tim Schakel<sup>a,\*</sup>, Boris Peltenburg<sup>a</sup>, Jan-Willem Dankbaar<sup>b</sup>, Carlos E. Cardenas<sup>c</sup>,  
Michalis Aristophanous<sup>c</sup>, Chris H.J. Terhaard<sup>a</sup>, Johannes M. Hoogduin<sup>b</sup>, Marielle E.P. Philippens<sup>a</sup>

<sup>a</sup> Department of Radiotherapy, University Medical Center, Utrecht, The Netherlands

<sup>b</sup> Department of Radiology, University Medical Center, Utrecht, The Netherlands

<sup>c</sup> Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, TX, USA

## ARTICLE INFO

## Keywords:

Radiotherapy  
Head and neck  
Target volume delineation  
Diffusion MRI  
PET

## ABSTRACT

**Background and purpose:** Diffusion weighted (DW) MRI may facilitate target volume delineation for head-and-neck (HN) radiation treatment planning. In this study we assessed the use of a dedicated, geometrically accurate, DW-MRI sequence for target volume delineation. The delineations were compared with semi-automatic segmentations on  $^{18}\text{F}$ -fluorodeoxyglucose (FDG) positron emission tomography (PET) images and evaluated for interobserver variation.

**Methods and materials:** Fifteen HN cancer patients underwent both DW-MRI and FDG-PET for RT treatment planning. Target delineation on DW-MRI was performed by three observers, while for PET a semi-automatic segmentation was performed using a Gaussian mixture model. For interobserver variation and intermodality variation, volumes, overlap metrics and Hausdorff distances were calculated from the delineations.

**Results:** The median volumes delineated by the three observers on DW-MRI were 10.8, 10.5 and 9.0 cm<sup>3</sup> respectively, and was larger than the median PET volume (8.0 cm<sup>3</sup>). The median conformity index of DW-MRI for interobserver variation was 0.73 (range 0.38–0.80). Compared to PET, the delineations on DW-MRI by the three observers showed a median dice similarity coefficient of 0.71, 0.69 and 0.72 respectively. The mean Hausdorff distance was small with median (range) distances between PET and DW-MRI of 2.3 (1.5–6.8), 2.5 (1.6–6.9) and 2.0 (1.35–7.6) mm respectively. Over all patients, the median 95th percentile distances were 6.0 (3.0–13.4), 6.6 (4.0–24.0) and 5.3 (3.4–26.0) mm.

**Conclusion:** Using a dedicated DW-MRI sequence, target volumes could be defined with good interobserver agreement and a good overlap with PET. Target volume delineation using DW-MRI is promising in head-and-neck radiotherapy, combined with other modalities, it can lead to more precise target volume delineation.

## 1. Introduction

Intensity modulated radiotherapy (RT) has allowed for conformal dose distributions, aiming to maximize the dose to the gross tumor volume (GTV) and minimize the dose to normal tissue. Accurate determination of the GTV is a key issue in radiotherapy [1] and forms the basis of treatment planning. However, it is also one of the major sources of uncertainty [2–4]. The accurate delineation of the GTV is a time consuming step, prone to errors. Delineations of the GTV using computed tomography (CT) and magnetic resonance imaging (MRI) show large variations between observers and an overestimation of the actual pathological tumor volume [5–9]. Additionally, FDG-PET has shown to

result in more conformal segmentations than CT or MRI [8,9].

To address the interobserver variation, automatic segmentation of the GTV would be ideal. Automatic segmentation is also preferable in online or adaptive MRI guided treatments. Several authors have proposed  $^{18}\text{F}$ -fluorodeoxyglucose (FDG) positron emission tomography (PET) as imaging modality for automatic segmentation [10–14] due to its high signal to background ratio. However, the use of PET for segmentation is challenging. The segmentations can depend on the contrast and noise characteristics of the PET images, which, apart from tumor characteristics, originate from different acquisition and reconstruction protocols [15,16]. Additionally, a variety of segmentation algorithms have been proposed which result in large variations in target

\* Corresponding author at: Department of Radiotherapy, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands.  
E-mail address: [t.schakel@umcutrecht.nl](mailto:t.schakel@umcutrecht.nl) (T. Schakel).

volumes [10,17]. Therefore, successful automatic segmentation requires a careful choice of acquisition and reconstruction parameters and segmentation algorithms. Other drawbacks are the increased radiation burden, resolution limitations and relatively high costs.

Diffusion weighted (DW) MRI might be an alternative imaging modality for FDG-PET as DW-MRI generates images with high contrast between tumor and surrounding normal tissue, making it a candidate to improve MRI based delineations and suitable for automatic segmentation. The contrast in DW-MRI is based on differences in the restriction of water diffusion on a microscopic level. The resulting high contrast images, along with the derived quantitative apparent diffusion coefficient (ADC) maps, could be used for delineation purposes [18,19]. However, the geometrically accurate acquisition of DW-MRI in the head-and-neck region is difficult due to the presence of air cavities causing large magnetic field inhomogeneities. Using conventional DW-MRI, acquired with echo planar imaging (EPI), image distortions up to centimeters can arise [20]. Recently, we have proposed the use of an alternative DW-MRI sequence; split acquisition of fast spin-echo (FSE) signal for diffusion imaging (SPLICE). Here, EPI is replaced with FSE for data acquisition, which leads to less distortions as it is relatively insensitive to magnetic field inhomogeneities introduced by the patient. The DW-SPLICE sequence has excellent geometric performance, allowing the DW-MRI images to be used for target delineation [21].

As the overestimation of the tumor volume results in a high burden on normal tissues surrounding the tumor due to the large treatment volume, the tumor has to be segmented as close (conformal) as possible. By definition the GTV only includes macroscopic tumor tissue, and requires a clinical target volume (CTV) margin to encompass microscopic spread. To avoid normal tissue burden, the CTV needs to be minimized. Recently, Ligtenberg et al. showed in an image validation study with pathology that GTVs automatically segmented from PET showed a tumor volume closest to the tumor volume on histology with a high coverage and the lowest overestimation [18]. Unfortunately, such validation studies cannot be performed in patients treated with primary radiotherapy. However, since the PET-based volume resembled the tumor volume best, and with the smallest overestimation, we chose to use PET-based segmentations as a reference to evaluate GTVs created using DW-SPLICE.

The aim of the present study was to assess the use of the dedicated DW-MRI sequence in target volume delineation. The GTVs were evaluated in terms of interobserver agreement, volume and spatial concordance with PET.

## 2. Methods and materials

### 2.1. Patients

A total of 15 head-and-neck patients were included in this study. The institutional review board provided a waiver of consent. These patients, scheduled for RT treatment, underwent a clinical FDG-PET/CT and an MRI exam, both in treatment position using a personalized head support and 5-point head-and-shoulder immobilization mask [22].

### 2.2. FDG-PET/CT imaging

Patients fastened six h before imaging on a mCT-Biograph PET/CT scanner (Siemens Medical Solutions, USA). An FDG dose of 2.0 MBq/kg was injected 60 min before image acquisition. Image acquisition was done at two bed positions, at four min per position. Image reconstruction was done using an ordered subset expectation maximization algorithm, utilizing point-spread function and time-of-flight information, at a reconstructed voxel size of  $2.04 \times 2.04 \times 1.5 \text{ mm}^3$  (AP  $\times$  RL  $\times$  FH) with body weight corrected standardized uptake values.

**Table 1**

Sequence parameters for the DW-SPLICE sequence. The protocol was adapted for large-scale imaging (denoted in round brackets).

Scan	DW-SPLICE
FOV (RL $\times$ AP $\times$ FH) [mm <sup>3</sup> ]	230 $\times$ 280 $\times$ 120 (200x200x120)
Acquired voxel size [mm <sup>2</sup> ]	1.8 $\times$ 1.8
Slice thickness [mm]	4
TE [ms]	52
TR [ms]	16,366 (15,844)
SENSE acceleration factor	2
(Echo) train length {dummies}	64 {1} (55 {1})
Bandwidth [Hz]	900.3 (858.5)
b-values [s/mm <sup>2</sup> ] {averages}	0 {2}, 800 {5}
Total acquisition time	4m38s (4m29s)

### 2.3. MR imaging

MR imaging was performed on a 3.0T Ingenia wide bore MR system (Philips Healthcare, The Netherlands) using 2-element Flex-M surface coils. The DW-SPLICE sequence was prototyped and loaded on the system, generating geometrically accurate diffusion weighted images [21]. Diffusion weighting was applied in an isotropic fashion using three orthogonal directions with b-values of 0 and 800 s/mm<sup>2</sup> (b0 and b800 respectively). ADC maps were generated on the MR system. Fat suppression was executed using spectral presaturation with inversion recovery. Detailed sequence parameters are given in Table 1. The exam also contained T<sub>1</sub> (with and without contrast) and T<sub>2</sub> weighted FSE scans.

### 2.4. Target volume delineations

For the acquired PET scans, the GTV was segmented semi-automatically using a Gaussian mixture model (GMM). With this method, voxels are classified based on statistical differences in their intensity distributions [23–25]. The GMM assumes independent observations, and models the image intensities with 10 Gaussian distributions. Following an initialization step by placing a selection box around the lesion, Gaussian distribution parameters are initialized automatically. The method was adapted for head-and-neck [24] by fully automating the initialization of the Gaussian distribution means as described in detail in previous publications [23,25]. Following initialization, the Gaussian distributions are fitted to the image intensities using maximum likelihood estimation method (MLEM). The algorithm then classifies regions to either tumor or background, generating binary masks.

The GTV was delineated in the diffusion weighted images by three observers. First the hyperintense region on the b800 image was identified, this was aided by a semi-automatic segmentation. On the b800 image, a seed point was placed centrally within the tumor and using a threshold of 50% of the maximum signal intensity, a region was segmented. Following this initial segmentation, the contours were manually adapted using the contrast of the ADC map, where the individual observers excluded regions of high ADC typically at the edge of the tumors. While doing this the b0 and b800 images were also available to the observers for reference. All contours (referred to as DW<sub>1</sub>, DW<sub>2</sub>, DW<sub>3</sub>) were converted to binary masks.

Additionally, as all of the patients received radiation treatment following imaging, the target volumes were retrieved from the treatment plans for these patients (GTV<sub>RT</sub>) as reference. The radiation oncologists mostly relied on CT and anatomical MRI (T<sub>1</sub> weighted, with and without contrast, and T<sub>2</sub> weighted), but PET and DW-MRI were also available.

### 2.5. Data analysis

All scans from the MRI exam were matched with the PET/CT using

rigid registration. The resulting transformations were used to resample both the PET and DW-SPLICE masks on the grid of the CT ( $0.98 \times 0.98 \times 2.0 \text{ mm}^3$ ), in order to compare the segmentations. While resampling the masks, the volume was kept the same on the new grid.

Using the resampled masks the delineated volumes for PET and DW-SPLICE were determined. For each observer the DW-MRI delineation was further compared to the PET segmentation by means of Dice similarity coefficients (DSC) and Hausdorff distance (HD) analysis. The DSC is a measure of spatial overlap and is defined as:  $DSC(A,B) = \frac{2(A \cap B)}{(A + B)}$  where A and B are the respective delineated regions and  $\cap$  is the intersection. A DSC of 0 indicates no spatial overlap at all and a DSC of 1 indicates complete overlap. The distance between two delineated contours is represented using the Hausdorff distance, which measures, for each point on a delineated contour, the Euclidean distance to the nearest point on the other contour. Per contour pair, the mean Hausdorff distance ( $HD_{\text{mean}}$ ) and the 95th percentile distance ( $HD_{95}$ ) were calculated. The results for each metric are given as the median (range) over all patients. We tested for significant differences ( $p < .05$ ) in metrics using a Wilcoxon signed rank test.

For the inter-observer analysis of the DW-SPLICE delineations, a generalized conformity index ( $CI_{\text{gen}} = \frac{\sum_{\text{pairs } ij} |A_i \cap A_j|}{\sum_{\text{pairs } ij} |A_i \cup A_j|}$ ) with  $A_i$  and  $A_j$  the respective delineated regions and  $\cup$  the union [26] over all observers was calculated per patient, because this metric is independent of the number of observers. Additionally, DSC were calculated per observer pair. Additionally, the HD analysis was also performed for the observer pairs as this metric is independent of the volume of the tumor and has clinical relevance as it is related to uncertainty margins.

### 3. Results

An example of the imaging data is given in Fig. 1, showing the correspondence between the different modalities as well as between the

## Volumes (logarithmic scale)

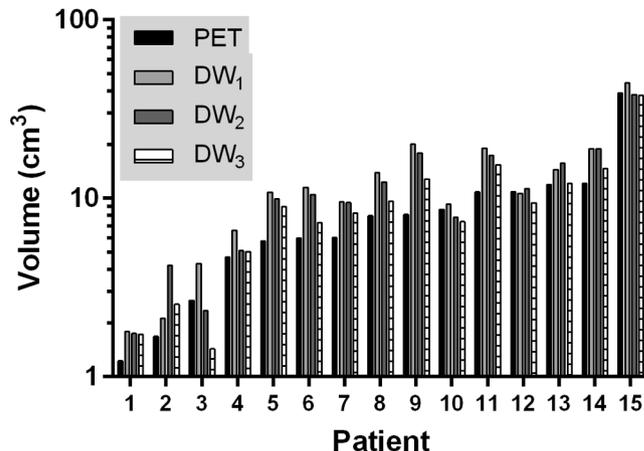


Fig. 2. The delineated volumes per patient for the different observers and modalities. Patients were ranked in ascending order of the volume determined on FDG-PET.

different observers on the diffusion weighted imaging. Additionally, anatomical T<sub>1</sub> weighted images after gadolinium contrast agent injection are shown for reference. Initial semi-automatic segmentation of the b800 images proved difficult in three out of fifteen patients, where the pixel intensity gradient was low around the tumor in images with limited SNR. For these patients, the tumor region was cropped manually to avoid other structures, such as unsuppressed fat, to be included in the region of interest.

#### 3.1. Volumes

The volumes found on the GMM segmentation of PET had a median volume of  $8.0 \text{ cm}^3$  (1.2–38.9), while the DW-SPLICE volumes were

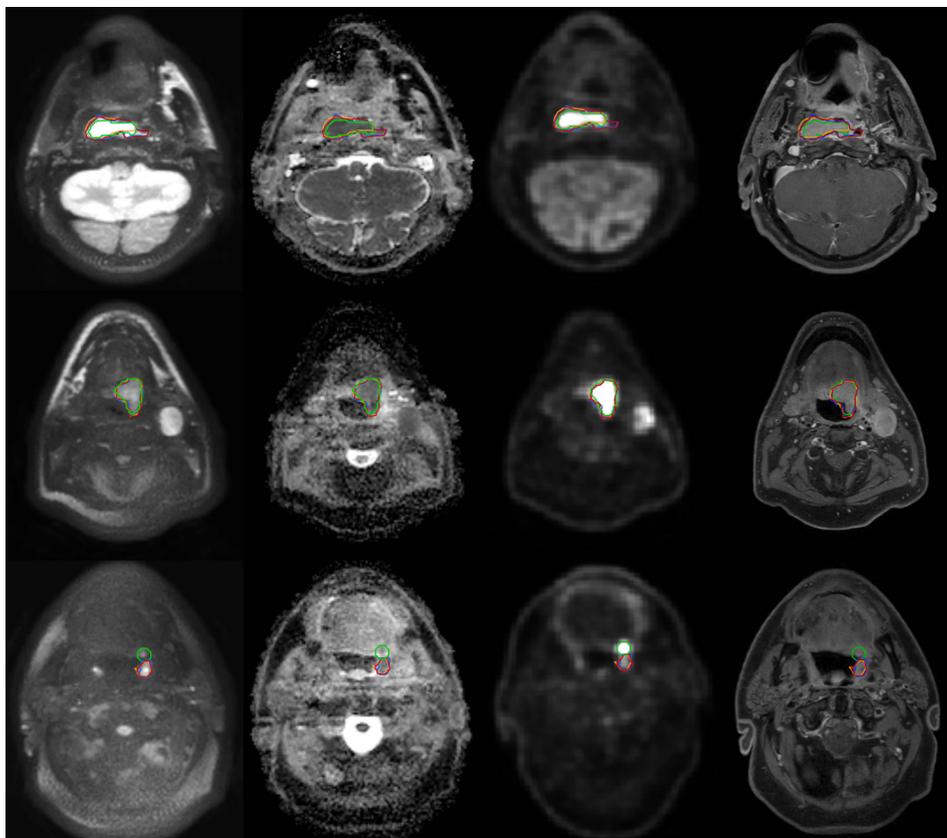


Fig. 1. Patient examples showing the correspondence of the delineated target volumes. Each row of images is from a single patient and shows transverse slices with, from left to right: DW b800; ADC map; FDG-PET; T1 weighted TSE mDixon water reconstruction after gadolinium contrast agent injection. The images are taken from patient 9, 12 and 1 respectively. The delineations from FDG-PET (green), DW<sub>1</sub> (red), DW<sub>2</sub> (blue) and DW<sub>3</sub> (orange) are shown on all imaging. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

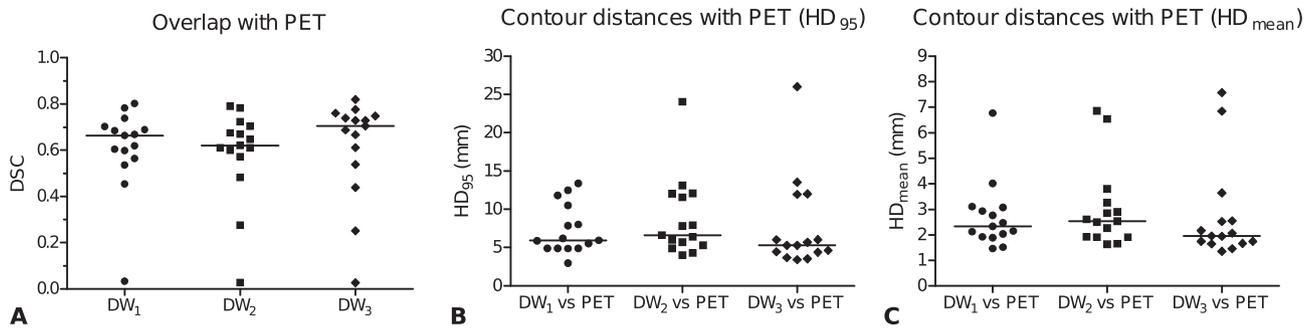


Fig. 3. The agreement between DW-MRI and FDG-PET as measured with Dice Similarity Coefficient (a), the 95th percentile contour distance (b) and the mean contour distance (c). The horizontal bars represent the median.

larger for all three observers: DW<sub>1</sub> 10.8 (1.8–44.3), DW<sub>2</sub> 10.5 (1.8–38.1), DW<sub>3</sub> 9.0 (1.4–37.8) cm<sup>3</sup> (Fig. 2). The difference with PET was significant for DW<sub>1</sub> ( $p < .05$ ) and DW<sub>2</sub> ( $p < .05$ ), but not for DW<sub>3</sub> ( $p > .05$ ). The median GTV<sub>RT</sub> was 15.6 cm<sup>3</sup>, a significant difference compared to both PET and all DW-MRI volumes. An overview of all delineated volumes is given in Supplementary Table 1.

3.2. Intermodality delineation overlap and distance analysis

The median (range) DSC of the PET segmentation with the each DW-MRI delineation was very similar for the three observers: 0.71 (0.03–0.80), 0.69 (0.03–0.79) and 0.72 (0.03–0.82) respectively (Fig. 3a). An overview of correspondence measures is given in Supplementary Table 1. The DSC was reasonable to good (0.44–0.82) in all but 2 patients; these were the patients with the smallest volumes. Patient 1 had practically no overlap between PET and DW-MRI. This patient showed two neighboring hotspots on PET, with the anterior one showing most tracer uptake so this region was segmented by the GMM method. On DW-MRI however, the posterior hotspot showed most diffusion restriction and low ADC (Fig. 1, bottom row). The other patient (number 2) showing poor correspondence, had a tumor located at the base of the tongue. The mismatch was mainly caused by tongue movement between the PET scan and MRI exam.

Using the same delineation pairs, distance analysis was performed (Fig. 3b and c) and both the HD<sub>95</sub> and HD<sub>mean</sub> were calculated. For all three observers the HD<sub>mean</sub> was small with median (range) distances between PET and DW-MRI of 2.3 (1.5–6.8), 2.5 (1.6–6.9) and 2.0 (1.35–7.6) mm respectively. Over all patients, the median 95th percentile distances were 6.0 (3.0–13.4), 6.6 (4.0–24.0) and 5.3 (3.4–26.0) mm. Similar to the DSC, patients 1 and 2 showed the largest distances between the PET and DW-MRI contours.

3.3. Interobserver agreement

The median CI<sub>gen</sub> over all patients was 0.73 (0.38–0.80) indicating good agreement among the different observers (Fig. 4a). This is also seen in the three pairwise DSCs, median (range): observer 1 and 2, 0.87 (0.52–0.90); observer 1 and 3, 0.83 (0.43–0.91); observer 2 and 3, 0.82

(0.59–0.87). Patient 2 and 3 showed the least interobserver agreement. These were patients with small tumor volumes, located in the oropharynx adjacent to lymphatic tissue (Waldeyer’s tonsillar ring).

Fig. 4b and c show the results of the distance analysis between observers. Generally, small distances, both HD<sub>95</sub> and HD<sub>mean</sub>, were found between the observer pairs except for the two patients (2 and 3) which also showed reduced CI<sub>gen</sub>.

4. Discussion

In this study we assessed semi-automated target volume delineation in head and neck cancer patients using a non-EPI diffusion weighted MRI sequence, DW-SPLICE. Furthermore, we investigated the variation between different observers on DW-MRI as well as the agreement between DW-MRI and <sup>18</sup>FDG-PET generated GTV delineations. The agreement between observers was good with a median CI<sub>gen</sub> of 0.73 and small distances between the individual contours (median HD<sub>95</sub> < 3.5 mm). The delineated volumes on DW-MRI were found to be significantly larger (median volumes 10.8 cm<sup>3</sup>, 10.5 cm<sup>3</sup> and 9.0 cm<sup>3</sup>) than the PET segmentations (median volume 8.0 cm<sup>3</sup>), nevertheless there was a substantial overlap between PET and DW-MRI with median DSCs of 0.71, 0.69 and 0.72 for the three different observers respectively. Both the PET and DW-MRI volumes were substantially smaller than the clinically used target volume, GTV<sub>RT</sub> (median volume 15.6 cm<sup>3</sup>), which is known to overestimate the GTV [8,9].

High b-value DW-MRI images have a similar appearance as PET images, which makes them relatively easy to interpret by the observers. This was demonstrated by the good agreement between the different observers in this study. The median CI<sub>gen</sub> over all patients was 0.73, where 0.7 is considered to be an indicator of good overlap [27]. The two cases showing a low CI<sub>gen</sub> concerned patients with small tumors located in the oropharynx, adjacent to lymphatic tissue. Lymphatic tissue shows diffusion restriction and thus has diffusion properties similar to that of tumors, i.e., showing high signal on high b-value DW-MRI images and a low ADC. Therefore, DW-MRI should always be used as an addition to high quality anatomical images. Other delineation studies in the head-and-neck area [28,29] show lower or similar CI<sub>gen</sub>

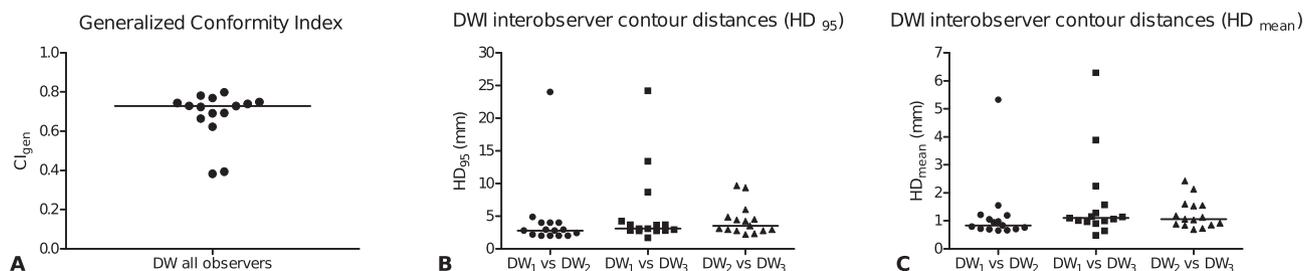


Fig. 4. Interobserver agreement for all patients on diffusion weighted imaging, expressed as generalized conformity index (a), contour distances HD<sub>95</sub> (b) and HD<sub>mean</sub> (c). The horizontal bars are the medians.

values. Jager et al. [5] showed 0.61 for laryngeal cancer using CT and 0.57 using CT/MR; Geets et al. reported 0.41 and 0.42 for laryngeal and oropharyngeal GTVs respectively [30]; Mukesh et al. reported 0.54 for CTV delineations on CT [31].

Despite the different biological background and acquisition method of DW-MRI and FDG-PET, there was a large overlap between PET and DW-MRI. For a large part both techniques identified the same target for treatment. Differences mainly occur at the edges of the delineated volumes, but the distances between the different contours were small in general. Over all patients, the median of the average distance between DW-MRI and PET contours was between 2.0 and 2.5 mm for the three different observers. For comparison, the acquired in-plane voxel sizes in DW-MRI and PET were 1.8 and 2.0 mm respectively. Furthermore, the PET GTVs, derived by the GMM segmentation, still require an appropriate CTV margin to encompass all microscopic tumor tissue. According to Ligtenberg et al. [8] this CTV margin is 5.2 mm in laryngeal and hypopharyngeal cancer. In a similar fashion, the DW-MRI GTV delineations will also require an appropriate CTV margin.

Generally, the difficulty of incorporating DW-MRI in the practice of target volume delineation in head-and-neck is the large geometrical distortion. To overcome this issue we used a dedicated sequence, DW-SPLICE, for the acquisition of diffusion images. Although this dealt with the geometrical distortions, some mismatch with other imaging modalities, like PET-CT can appear. Despite the fact that patients were fixated in an immobilization mask, some misalignment may remain since patients were scanned at two different time points. Additionally, internal motion of mobile structures due to breathing, tongue movement or swallowing can still occur. This could explain some of the differences in the delineations between DW-MRI and PET, especially for patient two.

The largest differences between DW-MRI and PET were found in small tumors, adjacent to lymphatic tissue. Both DW-MRI and FDG-PET are prone to false positives in these tissues, as it can show restricted diffusion behavior and increased FDG uptake. PET and DW-MRI visualize different biological processes and previous studies have investigated the relation between PET and DW-MRI in head-and-neck cancer [32–38]. However, these studies mainly looked at the relationship between SUV and ADC characteristics within the tumor and relating these to staging [37], histopathological parameters [35] or response prediction [33] and evaluation [34,38]. Houweling et al. [32] investigated a radiotherapy application of PET and DW-EPI after deformable registration. They specifically looked into targets for dose painting within a previously defined GTV. The spatial resolution is a limitation of both techniques. The detection limit and point spread function in PET could account for some of the differences found in the smaller lesions (i.e., patients 1 and 2). In PET the spatial resolution is mainly limited by the detector width, where in DW-MRI the resolution is limited due to the usage of a single-shot readout sequence. Advances in MRI sequence design (i.e., robust multi-shot DW-MRI sequences, faster imaging etc), could still provide additional gains in spatial resolution.

A limitation of the study is that no histopathology was available to compare the target volume with. In order to fully assess the value of DW-MRI for target volume delineation, it would be ideal to use pathology specimens in image validation studies [24–26]. Unfortunately, this is very challenging for some anatomical sites such as the oro- and nasopharynx. As we know from previous studies, GTV on CT and MRI overestimates the tumor largely and shows large interobserver variation [5–8]. Therefore, we compared the target volume with an automatic segmentation on PET, as the PET volume approaches the true tumor volume closest and the automatic segmentation method is validated versus pathology. The acquisition of only 2b-values,  $b = 0 \text{ s/mm}^2$  and  $b = 800 \text{ s/mm}^2$ , in the diffusion sequence is another limitation. This could result in an overestimation of the ADC value in areas with a large blood plasma volume [39]. Finally, the number of patients in this study is limited.

The target volumes derived from DW-MRI were generally smaller than the GTV defined in current clinical practice ( $DW_1$ ,  $DW_2$ ,  $DW_3$  vs  $GTV_{RT}$ ). These smaller target volumes already lead to smaller treatment volumes. The addition of DW-SPLICE to the current practice of target volume delineation can help reduce variation among observers, since it allows for semi-automatic segmentations. The region identified by both DW-MRI and PET could be used as a first estimate for the tumor outline. Subsequently, using conventional T1- and T2-weighted MRI and CT, the GTV delineation can be refined. The combination of all these imaging modalities can further improve target volume delineation.

In conclusion, using an optimized DW-MRI sequence, target volumes were defined with good interobserver agreement and a good overlap with PET. Target volume delineation using undistorted DW-MRI is promising in head-and-neck radiotherapy.

### Conflict of interest

The authors declare to have no conflict of interest.

### Acknowledgements

The authors thank Dr. Patricia Doornaert for helpful discussions.

Financial support for this work was provided by the Dutch Cancer Society (project UU 2011-5216).

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.phro.2017.12.004>.

### References

- [1] Baumann M, Krause M, Overgaard J, et al. Radiation oncology in the era of precision medicine. *Nat Rev Cancer* 2016;16:234–49. <http://dx.doi.org/10.1038/nrc.2016.18>.
- [2] Njeh CF. Tumor delineation: the weakest link in the search for accuracy in radiotherapy. *J Med Phys* 2008;33:136–40. <http://dx.doi.org/10.4103/0971-6203.44472>.
- [3] Weiss E, Hess CF. The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy theoretical aspects and practical experiences. *Strahlenther Onkol* 2003;179:21–30. <http://dx.doi.org/10.1007/s00066-003-0976-5>.
- [4] 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;123:412–8. <http://dx.doi.org/10.1016/j.radonc.2017.04.015>.
- [5] 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;10:26. <http://dx.doi.org/10.1186/s13014-014-0321-4>.
- [6] Caldas-Magalhaes J, Kooij N, Ligtenberg H, et al. The accuracy of target delineation in laryngeal and hypopharyngeal cancer. *Acta Oncol* 2015;54:1181–7. <http://dx.doi.org/10.3109/0284186X.2015.1006401>.
- [7] Cooper JS, Mukherji SK, Toledano AY, et al. An evaluation of the variability of tumor-shape definition derived by experienced observers from CT images of supraglottic carcinomas (ACRIN protocol 6658). *Int J Radiat Oncol Biol Phys* 2007;67:972–5. <http://dx.doi.org/10.1016/j.ijrobp.2006.10.029>.
- [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;123:63–70. <http://dx.doi.org/10.1016/j.radonc.2017.02.005>.
- [9] Daisne JF, Duprez T, Weynand 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. <http://dx.doi.org/10.1148/radiol.2331030660>.
- [10] Zaidi H, El Naqa I. PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques. *Eur J Nucl Med Mol Imaging* 2010;37:2165–87. <http://dx.doi.org/10.1007/s00259-010-1423-3>.
- [11] Geets X, Lee JA, Bol A, Lonnewux M, Gregoire V. A gradient-based method for segmenting FDG-PET images: methodology and validation. *Eur J Nucl Med Mol Imaging* 2007;34:1427–38. <http://dx.doi.org/10.1007/s00259-006-0363-4>.
- [12] Zaidi H, Abdoli M, Fuentes CL, El Naqa IM. Comparative methods for PET image segmentation in pharyngolaryngeal squamous cell carcinoma. *Eur J Nucl Med Mol Imaging* 2012;39:881–91. <http://dx.doi.org/10.1007/s00259-011-2053-0>.
- [13] Cheebsumon P, Boellaard R, de Ruyscher D, et al. Assessment of tumour size in PET/CT lung cancer studies: PET- and CT-based methods compared to pathology. *EJNMMI Res* 2012;2:56. <http://dx.doi.org/10.1186/2191-219X-2-56>.
- [14] Daisne JF, Sibomana M, Bol A, et al. Tri-dimensional automatic segmentation of

- PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms. *Radiother Oncol* 2003;69:247–50.
- [15] Brambilla M, Matheoud R, Secco C, et al. Threshold segmentation for PET target volume delineation in radiation treatment planning: The role of target-to-background ratio and target size. *Med Phys* 2008;35:1207–13. <http://dx.doi.org/10.1118/1.2870215>.
- [16] Matheoud R, Monica PD, Loi G, et al. Influence of reconstruction settings on the performance of adaptive thresholding algorithms for FDG-PET image segmentation in radiotherapy planning. *J Appl Clin Med Phys* 2011;12:115–32. <http://dx.doi.org/10.1120/jacmp.v12i2.3363>.
- [17] Schinagl DAX, Vogel WV, Hoffmann AL, et al. Comparison of Five Segmentation Tools for 18F-Fluoro-Deoxy-Glucose Positron Emission Tomography Based Target Volume Definition in Head and Neck Cancer. *Int J Radiat Oncol Biol Phys* 2007;69:1282–9. <http://dx.doi.org/10.1016/j.ijrobp.2007.07.2333>.
- [18] Burbach JP, Kleijnen JP, 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. <http://dx.doi.org/10.1016/j.radonc.2015.10.030>.
- [19] van Rossum PS, van Lier AL, van Vulpen M, et al. Diffusion-weighted magnetic resonance imaging for the prediction of pathologic response to neoadjuvant chemoradiotherapy in esophageal cancer. *Radiother Oncol* 2015;115:163–70. <http://dx.doi.org/10.1016/j.radonc.2015.04.027>.
- [20] Schakel T, Hoogduin JM, Terhaard CH, Philippens ME. Diffusion weighted MRI in head-and-neck cancer: geometrical accuracy. *Radiother Oncol* 2013;109:394–7. <http://dx.doi.org/10.1016/j.radonc.2013.10.004>.
- [21] Schakel T, Hoogduin JM, Terhaard CH, Philippens MEP. Technical Note: diffusion-weighted MRI with minimal distortion in head-and-neck radiotherapy using a turbo spin echo acquisition method. *Med Phys* 2017. <http://dx.doi.org/10.1002/mp.12363>.
- [22] Verduijn GM, Bartels LW, Raaijmakers CP, et al. Magnetic resonance imaging protocol optimization for delineation of gross tumor volume in hypopharyngeal and laryngeal tumors. *Int J Radiat Oncol Biol Phys* 2009;74:630–6. <http://dx.doi.org/10.1016/j.ijrobp.2009.01.014>.
- [23] Aristophanous M, Penney BC, Martel MK, Pelizzari CA. A Gaussian mixture model for definition of lung tumor volumes in positron emission tomography. *Med Phys* 2007;34:4223–35. <http://dx.doi.org/10.1118/1.2791035>.
- [24] Caldas Magalhaes J, Raaijmakers CP, Aristophanous M, et al. FDG-PET semi automatic segmentation methods for GTV delineation in laryngeal and hypopharyngeal cancer. *Int J Radiat Oncol Biol Phys* 2014;90:S536. <http://dx.doi.org/10.1016/j.ijrobp.2014.05.1631>.
- [25] Aristophanous M, Pelizzari CA. The evaluation of a highly automated mixture model based technique for PET tumor volume segmentation *Medical Imaging: SPIE*; 2008. p. 12.
- [26] 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. <http://dx.doi.org/10.1088/0031-9155/54/9/018>.
- [27] Zou KH, Warfield SK, Bharatha A, et al. Statistical validation of image segmentation quality based on a spatial overlap index: scientific reports. *Acad Radiol* 2004;11:178–89. [http://dx.doi.org/10.1016/S1076-6332\(03\)00671-8](http://dx.doi.org/10.1016/S1076-6332(03)00671-8).
- [28] Loo SW, Martin WMC, Smith P, Cherian S, Roques TW. Interobserver variation in parotid gland delineation: a study of its impact on intensity-modulated radiotherapy solutions with a systematic review of the literature. *Brit J Radiol* 2012;85:1070–7. <http://dx.doi.org/10.1259/bjr/32038456>.
- [29] 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;83:220–7. <http://dx.doi.org/10.1016/j.ijrobp.2011.05.060>.
- [30] Geets X, Daisne J-F, Arcangeli S, et al. Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: Comparison between CT-scan and MRI. *Radiother Oncol* 2005;77:25–31. <http://dx.doi.org/10.1016/j.radonc.2005.04.010>.
- [31] Mukesh M, Benson R, Jena R, et al. Interobserver variation in clinical target volume and organs at risk segmentation in post-parotidectomy radiotherapy: can segmentation protocols help? *Brit J Radiol* 2012;85:e530–6. <http://dx.doi.org/10.1259/bjr/66693547>.
- [32] Houweling AC, Wolf AL, Vogel WV, et al. FDG-PET and diffusion-weighted MRI in head-and-neck cancer patients: implications for dose painting. *Radiother Oncol* 2013;106:250–4. <http://dx.doi.org/10.1016/j.radonc.2013.01.003>.
- [33] Schouten CS, de Bree R, van der Putten L, et al. Diffusion-weighted EPI- and HASTE-MRI and 18F-FDG-PET-CT early during chemoradiotherapy in advanced head and neck cancer. *Quant Imaging Med Surg* 2014;4:239–50. <http://dx.doi.org/10.3978/j.issn.2223-4292.2014.07.15>.
- [34] Schouten CS, de Graaf P, Alberts FM, et al. Response evaluation after chemoradiotherapy for advanced nodal disease in head and neck cancer using diffusion-weighted MRI and 18F-FDG-PET-CT. *Oral Oncol* 2015;51:541–7. <http://dx.doi.org/10.1016/j.oraloncology.2015.01.017>.
- [35] Surov A, Stumpp P, Meyer HJ, et al. Simultaneous (18)F-FDG-PET/MRI: Associations between diffusion, glucose metabolism and histopathological parameters in patients with head and neck squamous cell carcinoma. *Oral Oncol* 2016;58:14–20. <http://dx.doi.org/10.1016/j.oraloncology.2016.04.009>.
- [36] Varoquaux A, Rager O, Lovblad KO, et al. Functional imaging of head and neck squamous cell carcinoma with diffusion-weighted MRI and FDG PET/CT: quantitative analysis of ADC and SUV. *Eur J Nucl Med Mol Imaging* 2013;40:842–52. <http://dx.doi.org/10.1007/s00259-013-2351-9>.
- [37] Fruehwald-Pallamar J, Czerny C, Mayerhoefer ME, et al. Functional imaging in head and neck squamous cell carcinoma: correlation of PET/CT and diffusion-weighted imaging at 3 Tesla. *Eur J Nucl Med Mol Imaging* 2011;38:1009–19. <http://dx.doi.org/10.1007/s00259-010-1718-4>.
- [38] Subesinghe M, Scarsbrook AF, Sourbron S, et al. Alterations in anatomic and functional imaging parameters with repeated FDG PET-CT and MRI during radiotherapy for head and neck cancer: a pilot study. *BMC Cancer* 2015;15:137. <http://dx.doi.org/10.1186/s12885-015-1154-8>.
- [39] Le Bihan D, Breton E, Lallemand D, et al. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. *Radiology* 1988;168:497–505. <http://dx.doi.org/10.1148/radiology.168.2.3393671>.