AJNR

Comparison of Diffusion-weighted MRI using Singe-Shot Echo-planar Imaging (SS-EPI) and Split Acquisition of Fast Spin Echo Signal (SPLICE) Imaging, a non-EPI technique, in Tumors of the Head and Neck

Hedda J. van der Hulst, Loes Braun, Bram Westerink, Georgios Agrotis, Leon C. ter Beek, Renaud Tissier, Milad Ahmadian, Roland M. Martens, Jan W. Casselman, Regina G.H. Beets-Tan, Michiel W.M. van den Brekel and Jonas A. Castelijns

AJNR Am J Neuroradiol published online 9 October 2024
http://www.ajnr.org/content/early/2024/10/09/ajnr.A8529

Published October 9, 2024 as 10.3174/ajnr.A8529

ORIGINAL RESEARCH

Comparison of Diffusion-weighted MRI using Singe-Shot Echo-planar Imaging (SS-EPI) and Split Acquisition of Fast Spin Echo Signal (SPLICE) Imaging, a non-EPI technique, in Tumors of the Head and Neck.

Hedda J. van der Hulst[1,2,3*], Loes Braun[1], Bram Westerink[1], Georgios Agrotis[1,4], Leon C. ter Beek[5], Renaud Tissier[6], Milad Ahmadian[1,3,7], Roland M. Martens[8], Jan W. Casselman[9,10], Regina G.H. Beets-Tan[1,2,11], Michiel W.M. van den Brekel[3,12], Jonas A. Castelijns[1]

ABSTRACT

BACKGROUND AND PURPOSE: Diffusion-weighted imaging (DWI) using single-shot echo planar imaging (DW-EPI) is susceptible to distortions around air-filled cavities and dental fillings, typical for the head and neck area. Non-EPI, Split acquisition of fast spin echo signals for diffusion imaging (DW-SPLICE) could reduce these distortions and enhance image quality, thereby potentially improving recurrence assessment in squamous cell carcinoma (SCC) of the head and neck region. This study evaluated whether DW-SPLICE is a viable alternative to DW-EPI through quantitative and qualitative analyses.

MATERIALS AND METHODS: The DW-SPLICE sequence was incorporated into the standard 3.0T head and neck MRI protocol with DW-EPI. Retrospective analysis was conducted on two subgroups: firstly benign or malignant lesions, and secondly post-treatment SCC recurrence. In both subgroups Image quality and distortion were scored by two independent radiologists, blinded for DW-technique, and evaluated using mixed-effect linear models. Lesion apparent diffusion coefficient (ADC) values were assessed with inter-class correlation (ICC) and Bland-Altman analyses. DWI's delineation geometric similarity to T1-weighted post-contrast (T1Wc) MRI was evaluated using the Dice Similarity Coefficient (DSC) before and after registration. Recurrence in post-treatment SCC scans was evaluated by the same two radiologists blinded for DW-technique. Recurrence detection rates were then compared between DW-SPLICE and DW-EPI using mixed logistic regression a**t six months and at** one-year post-scan follow-up data.

RESULTS: From August 2020 to January 2022, 55 benign or malignant lesion scans (55 patients) and 74 post-treatment SCC scans (66 patients) were analyzed. DW-SPLICE scored better on image quality and showed less overall distortion than DW-EPI ($0.04<p<0.001$). There was high ADC measurement reliability (ICC=0.93, $p<0.001$), though a proportional bias was also observed ($\beta=0.11$, $p=0.03$), indicating the bias increases as ADC values rise. DW-SPLICE exhibited greater geometric similarity to T1Wc before registration (DSC 0.63 vs 0.47, $p<0.001$) and outperformed DW-EPI by more accurately identifying recurrences after one year (OR=0.96, $p=0.05$) but not after six months (OR=0.72, $p=0.13$).

CONCLUSIONS: DW-SPLICE surpasses DW-EPI on image distortion and quality and improves diagnostic reliability for detecting recurrent or residual SCC on 3T MRI of the HN. Consistent use of one method for follow-up is advised, as ADC values are not completely interchangeable. Integrating DW-SPLICE can significantly improve tumor assessments in clinical practice.

ABBREVIATIONS: ANTs = Advanced Normalization Tools; DSC = Dice Similarity Coefficient; DW-EPI = Diffusion-weighted single-shot echo planar imaging; DW-MS-EPI = Diffusion-weighted multi-shot echo planar imaging; DW-SPLICE = Diffusion-weighted split acquisition of fast spin echo signals for diffusion imaging; DW-TSE = Diffusion-weighted Turbo Spin Echo; ICC = Intraclass Correlation Coefficient; ROC = Receiver Operating Characteristic; SCC = Squamous cell carcinoma; T1WIc = T1 weighted imaging with gadolinium contrast.

Received month day, year; accepted after revision month day, year.
[1]Department of Radiology, The Netherlands Cancer Institute, Amsterdam, The Netherlands.
[2]GROW School for Oncology and Developmental Biology, University of Maastricht, Maastricht, The Netherlands.
[3]Department of Head and Neck Oncology and Surgery, The Netherlands Cancer Institute, Amsterdam, The Netherlands.
[4]Department of Radiology, University Hospital of Larissa, Thessaly, Greece
[5]Department of Clinical Physics, The Netherlands Cancer Institute, Amsterdam, The Netherlands
[6]Biostatistics Unit, The Netherlands Cancer Institute, Amsterdam, The Netherlands.
[7]Amsterdam Center for Language and Communication, University of Amsterdam, Amsterdam, The Netherlands.
[8]Department of Radiology and Nuclear Medicine, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands.
[9]Dalhousie University, Department of Radiology, Halifax, NS, Canada.
[10]Department of Radiology, AZ St Lucas Gent, Ghent, Belgium.
[11]Department of Regional Health Research, University of Southern Denmark, Odense, Denmark.
[12]Department of Oral and Maxillofacial Surgery, Amsterdam University Medical Center, Amsterdam, The Netherlands.

The authors declare no conflicts of interest related to the content of this article.

Please address correspondence to Hedda J. van der Hulst, MD, The Netherlands Cancer Institute, Department Radiology, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands; h.vd.hulst@nki.nl.

SUMMARY SECTION

PREVIOUS LITERATURE: Diffusion-weighted imaging (DWI) using single-shot echo planar imaging (DW-EPI) is known for distortions around air-filled cavities and dental fillings in the head and neck region. Non-EPI, Split acquisition of fast spin echo signals for diffusion imaging (DW-SPLICE) was proposed to reduce these distortions and improve image quality. Previous studies suggested potential benefits but lacked comprehensive evaluation.

KEY FINDINGS: DW-SPLICE showed superior image quality, less distortion, and higher geometric similarity to T1Wc MRI than DW-EPI at 3.0T. DW-SPLICE also more accurately identified recurrences after one year but not after six months.

KNOWLEDGE ADVANCEMENT: This study demonstrates that DW-SPLICE significantly enhances diagnostic reliability for detecting recurrent or residual SCC in head and neck MRI, suggesting its potential for integration into clinical practice at 3.0T machines, though more research is needed at other field strengths.

Copyright 2024 by American Society of Neuroradiology.

INTRODUCTION

Diffusion-weighted imaging (DWI), as a functional sequence of magnetic resonance imaging (MRI), excels in differentiating post-treatment effects from tumor recurrence compared to anatomical MR-imaging.[1] DWI's sensitivity to local water mobility on a microstructural level allows for it to be less affected by post-radiotherapy effects and inflammation, making it irreplaceable for finding and monitoring cancers such as squamous cell carcinoma (SCC) in the head and neck (HN) region.[2,3]

DWI in oncological HN-protocols is traditionally performed using the single-shot echo planar imaging (DW-EPI) technique due to its relatively fast read-out and good signal-to-noise ratios (SNR). However, DW-EPI is highly sensitive to susceptibility variations, mainly arising from air-tissue interfaces and metal, leading to significant artefacts, such as geometric distortions, and signal voids and accumulations.[4,5] This is especially a problem for the head and neck area, considering the air-filled cavities, dental fillings and reconstruction materials in this anatomical region.

To address these challenges, alternative DWI techniques like multi-shot echo planar imaging (DW-MS-EPI) and Turbo Spin Echo (DW-TSE) have been explored.[4]

DW-TSE, known for its efficacy in detecting middle ear cholesteatoma[6], has shown promise in head and neck oncology by differentiating between benign and malignant lesions based on the apparent diffusion coefficient (ADC).[7-11] However, adverse results have also been reported, indicating that DW-TSE has limited value in predicting malignancy.[12] Given that many studies on DW-TSE have been conducted at suboptimal conditions such as a low magnetic field strength (1.5T)[7,9,11,12] or small sample sizes[4,7,8,11], and have not yet evaluated its clinical effectiveness in detecting HNSCC recurrence, further investigation is warranted.

DWI scans in general have to be scanned under Carr Purcell Meiboom Gill (CPMG) conditions, dictating the phase relations and timings in the sequence. Incorporating the diffusion gradients for a DW-TSE in combination with the longer scanning time compared to DW-EPI, the chance of CPMG conditions being violated due to subject motion is high.[4] Motion causes phase errors and unstable echo trains and can lead to signal loss or even signal voids in DW-TSE.[5] To mitigate this, modified TSE-methods like split acquisition of fast spin echo signal for diffusion imaging (SPLICE) have been introduced.[13] DW-SPLICE separates the interfering echo contributions by independently acquiring and reconstructing each echo parity and afterward their signal magnitudes are combined. This technique offers less sensitivity to susceptibility and motion artefacts, albeit with a trade-off of a lower SNR compared to DW-EPI.[4,13] Sequence diagram of the DW-SPLICE and the phase behavior of several echo pathways have been previously visualized by Schick et al.[13] This alternative TSE-based technique has shown its potential for the abdominal and head and neck region MRI and MR-LINAC for improving DWI quality by mitigating artifacts.[4,10,11,14]

The aim of this study is to evaluate whether DW-SPLICE is a better choice than DW-EPI for lesion and recurrence assessment in the head and neck region. The analyses will focus on quantitative differences in measured ADC-values and the geometric similarity of both techniques compared to anatomical MRI. We also qualitatively evaluate differences through scoring image quality and distortion. Additionally, this study will examine the use of DW-EPI and DW-SPLICE to identify residual or recurrent tumor in post-treatment HNSCC scans, comparing results with one year of clinical follow-up data.

MATERIALS AND METHODS

2.1 Population and subgroups:

In this single-center study at the Netherlands National Cancer Institute (NKI-AVL), head and neck MRI scans performed on two 3.0T MRI-scanners between August 2020 to January 2022 were analyzed. An additional DW-SPLICE sequence was added to the standard head and neck protocol in the context of protocol enhancement. The additional DW-SPLICE scan was not performed in case of accumulating waiting times in the clinics or when a longer table-time would be too high of a burden to the patient, as determined by the radiographers. Patients were not preselected on clinical parameters before imaging. Patients were eligible for consecutive inclusion if the additional DW-SPLICE sequence was conducted in addition to the conventional DW-EPI scan.

To evaluate the different aspects of the DWI techniques, two subgroups of MRI scans were retrospectively selected for analysis. The first subgroup, the so-called lesion-assessment subgroup, included MRI scans of both datasets of untreated benign or malignant lesions or MRI scans of a recurrent lesion after stand-alone surgery. Lesions needed to be visible on at least a b1000 image or ADC map of both the DW-EPI and the DW-SPLICE scan. A minimal lesion volume of 0.5 cm3 was required for analyses. The second subgroup, the so-called SCC recurrence assessment subgroup, included post-treatment MRI scans of SCC in the head and neck region. All applicable post-treatment scans were included, regardless of the time between the imaging and the treatment or treatment type.

The available follow-up data of the patients was organized in accordance with the standard clinical check-up schedule consisting of clinical examinations including a flexible laryngoscopy every three months, additional scans or biopsies were acquired when indicated. Actual recurrence is measured at six months and one year after the assessed MRI scan was conducted.

The Standards for Reporting Diagnostic accuracy studies (STARD)[15] methodology has been followed where applicable. An overview of the STARD checklist and the application can be found in Supplementary Material Appendix 1.

2.2 MR Imaging Acquisition:

The MR-examinations were acquired on a 3.0T Achieve dStream and a 3.0T Ingenia system from the same vendor (Philips Healthcare, Best, the Netherlands, software version 5.6.1) and conducted with a dedicated 20 channel head and neck coil. Aside from the DWI scans, the standard head and neck protocol consisted of a Short-TI Inversion Recovery (STIR), a T1 weighted scan with and without gadolinium contrast agent (T1WI and T1WIc) and an isotropic 3D T1 weighted scan post contrast. The DW-EPI had a TE of 67.2-75.8 msec, a TR of 2840.8-6525.9 msec over the two scanners, with 4.0mm slice thickness, and scan time ranging from 110.8-254.5 seconds. The DW-SPLICE had a TE/TR of 62.4-71.6 msec / 4250.4-5659.3 msec over the two scanners, 4.0mm slice thickness, and a scan time ranging from 187.0-249.0 seconds. b-value of 0, 200, 1000 s/mm2 were collected for both DWI-techniques. ADC maps were calculated using all available b-values. Full imaging parameters of the DWI's and T1WIc can be found in Table 1.

Table 1: MRI Scanning Parameters Overview. Detailed MRI scanning parameters for the DW-EPI, DW-SPLICE and T1WIc sequences of all included patients (N = 121) across the two different MRI scanners.
*TR: repetition time, TE: echo time, N: total number of patients, n: number of patients in subcohort, *only lesion-assessment subgroup values.*

	DW – EPI	DW – SPLICE	T1WIc*

Scanner	Achieva dStream	Ingenia	Achieva dStream	Ingenia	Achieva dStream	Ingenia
Nr. of scans	n = 94	n = 35	n = 94	n = 35	n = 25*	n = 9*
Field strength	3T	3T	3T	3T	3T	3T
b-values	0, 200, 1000	0, 200, 1000	0, 200, 1000	0, 200, 1000	-	-
Voxel size (mm)	0.9 x 0.9	0.9 x 0.9	0.9 x 0.9	0.9 x 0.9	0.5 x 0.5	0.5 x 0.5
Slice thickness (mm)	4.0	4.0	4.0	4.0	3.0	3.0
TR (msec)	3603.8-5016.2	2840.8-6525.9	4344.3-5659.3	4250.4-4781.6	692.5-873.4	718.3-842.1
TE (msec)	67.2-68.7	75.3-75.8	62.4-62.9	71.0-71.6	10-16.7	10
Duration (s)	140.6-195.6	110.8-254.5	191.1-249.0	187.0-210.4	45.4-159.5	118.5-268.7
Parallel acquisition technique	SENSE	SENSE	SENSE	SENSE	Compressed SENSE	Compressed SENSE

2.3 Data preparation:

The DW-EPI and DW-SPLICE scans of each patient were separated, randomized, and blinded for DWI technique and patient-specific information within their subgroup, creating double the amount of cases compared to the amount of patients. For every DWI case the b0, b1000, ADC and T1WIc were available for the readers. Within the lesion-assessment subgroup, the T1WIc scans received separate randomization and blinding for the purposes of delineation, with readers additionally having access to corresponding STIR images.

2.4 Delineation and registration:

All scans in the lesion-assessment subgroup were delineated independently on the blinded DWI b1000 scans and on the T1WIc scans by a physician researcher (H.H.) and supervised by a radiologist (G.A.) with 5 years of radiological experience. Regions of interest (ROIs) were manually placed on the entire tumor volume using 3D-Slicer software (version 4.10.2; http://www.slicer.org). ROIs were placed in accordance with areas of high signal intensity at b1000, or, when not discernible at b1000, high or low intensity areas on the ADC map. When possible, ROIs are chosen for their high signal intensity on b1000 coupled with low signal intensity on ADC, to depict diffusion restriction.[16] Large cystic or necrotic areas were excluded if additional, non-cystic or non-necrotic, tissue could be identified, ensuring these areas did not impact the ADC measurement. An exception was applied to lesions that were exclusively cystic in nature. ROIs delineated on the DWI b1000 were used for the extraction of the ADC values. Raters were blinded for lesion type during delineation.

To determine the geometric similarity of the DWI delineations to anatomical MRI, DW-EPI and DW-SPLICE scans were registered to their corresponding T1WIc scan using Advanced Normalization Tools (ANTs) (version 0.0.7; http://stnava.github.io/ANTs/). After registration, the same transformation is applied to the DWI delineations. The volume of the DW-EPI and DW-SPLICE delineation that overlaps with the T1WIc delineation is measured before and after registration.

2.5 Qualitative evaluation:

All data from both subgroups were independently scored on image quality and distortion by two head and neck radiologist (L.B. and B.W.) with 5 and 3 years of experience, using a standardized score form. Image quality was defined as: 1= non-diagnostic quality, 2= poor quality, 3= acceptable quality, 4= good quality, 5= excellent quality.[17] Distortion was scored on the overall image and also specifically on the lesion location for all cases within the lesion-assessment subgroup. Distortion was defined as: 0= no image distortion, 1= mild image distortion , 2= moderate image distortion, 3= severe image distortion.[18] Cases within the SCC recurrence assessment subgroup were additionally scored for presence of recurrent or residual disease using a binary scoring system, defined as: 0= not suspect for recurrence or residual disease and 1= suspect for recurrence or residual disease. Consistency in scoring among raters was ensured through a kick-off meeting, followed by an evaluation after the first five cases. Full score forms can be found in Supplementary Material Appendix 2.

2.6 Statistical analyses:

All analyses were done using R statistics (version 4.3.3). A p-value < 0.05 was considered statistically significant.

The association between the qualitatively scored Image quality and distortion of both the DWI type (EPI or SPLICE) was assessed using a linear mixed effect model. This model was used to incorporate the dependence existing between observations in the dataset as a result of the two independent raters and the potential scoring link between the DWI images of the same case. Therefore, two random intercepts were used in the model at the two fixed rater level (two radiologists) and the MRI scan pair level (DW-SPLICE and DW-EPI). Additionally, the Inter-rater variability is assessed using a quadratically weighted Cohens kappa.

ADC value differences between DW-EPI and DW-SPLICE were analyzed using the Type 3 Intraclass Correlation Coefficient (ICC) for agreement and consistency, and Bland-Altman regression to detect any systematic bias or significant discrepancies. The geometric similarity of the DWI relative to the T1WIc was evaluated using the Dice Similarity Coefficient (DSC), Recall, and Precision score. The DSC evaluates how closely the areas delineated on DWI and T1WIc overlap. The Recall score focusses on the effectiveness of identifying T1WIc areas by measuring the proportion of T1WIc delineation correctly captured within the overlapping area with DWI delineation. The Precision score evaluates the precision of DWI delineation by quantifying the proportion of DWI delineation that does not overlap with T1WIc delineation. Comparisons of the scores between the DWI-techniques were made using Wilcoxon signed rank tests. Formulas of the DSC, recall and precision score are defined in Supplementary Material Appendix 3.

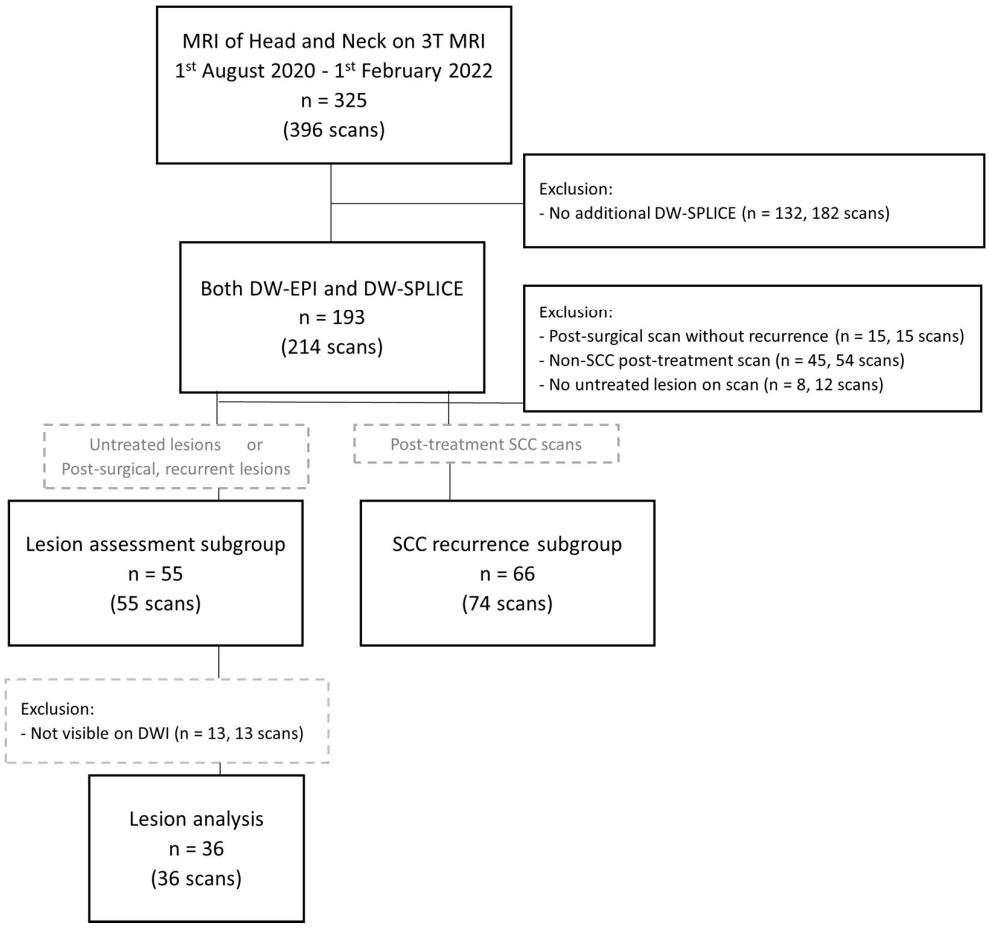
In the SCC recurrence subgroup, the Inter-rater variability is assessed using Cohens kappa. The recurrence score is assessed using mixed logistic regression to adjust for the inter- intrareader variability, using the binary recurrence score and the actual outcome data after both six months and after one year post-scan follow-up. The model uses the agreement between the scorers and the actual outcome. The random intercepts applied in this analyses were the same as for the image quality and distortion analyses, the two fixed rater level (two radiologists) and the MRI scan pair level (DW-SPLICE and DW-EPI). Additionally, the diagnostic accuracy and receiver-operating characteristic curve (ROC) were calculated after both six months and after one year post-scan follow-up as a supplementary test. While less suitable to the data due to the lack of incorporating the potential intercepts, these tests provide an additional insight into the data.

RESULTS

Between August 2020 and January 2022, a total of 396 MR scans targeting the head and neck area were completed on the selected 3.0T scanners in the NKI-AVL. In 214 of these scans the additional DW-SPLICE sequence was added to the scan protocol. 55 of these 214 scans depicted untreated benign or malignant lesions of the head and neck or recurrent masses after stand-alone surgical treatment and were included in the lesion-assessment subgroup.

Additionally, 74 of these 214 scans depicted post-treatment SCC of the head and neck region. These scans were selected for the SCC recurrence subgroup. See Figure 1 for the full flow diagram and exclusion criteria. Baseline characteristics of all subgroups can be found in Supplementary Material Appendix 4.

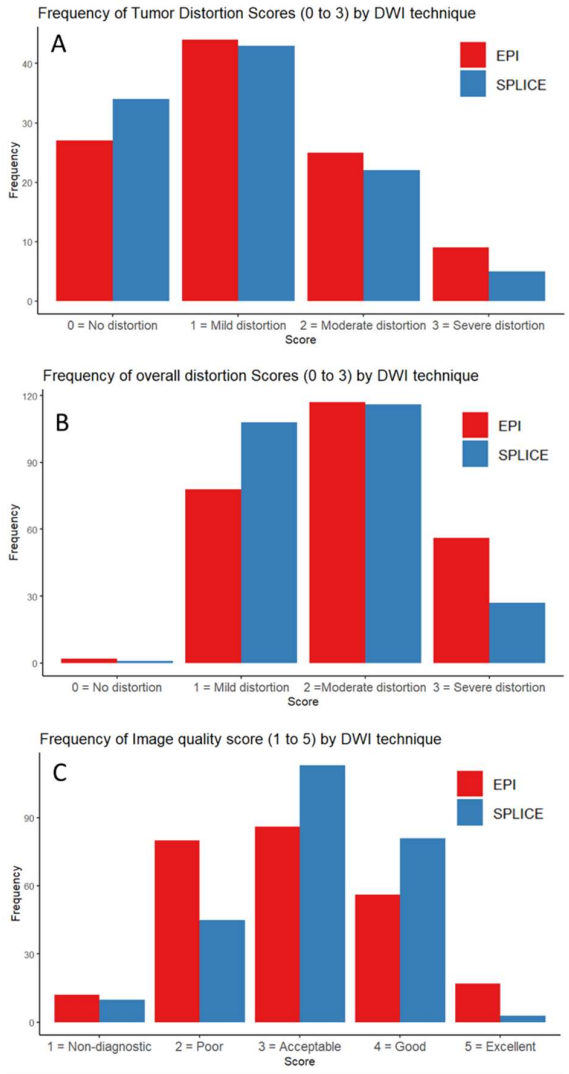
Figure 1. Patient inclusion and flow diagram. Figure illustrating the patient inclusion process (overall N = 121 (Lesion assessment subgroup n = 55, SCC recurrence subgroup n = 66)), detailing reasons for exclusion, and outlining the inclusion criteria and sample sizes for the separately analyzed subgroups. *n: number of patients in subcohort*



3.1 Image quality and distortion assessment:

All 129 scans from 121 patients in the lesion and SCC recurrence subgroups were scored for image quality and distortion. Baseline characteristics can be found in Supplementary Material Appendix 4. Total frequencies of scores on the (estimated) tumor distortion, the overall image distortion and on image quality are depicted in Figure 2. DW-SPLICE scans are more frequently rated as having "no to mild distortion" compared to DW-EPI scans, both for tumor-specific (74.0% vs. 67.6%) and overall distortion (43.3% vs. 31.6%). In contrast, DW-EPI scans more often receive "moderate to severe distortion" ratings (tumor-specific: 32.3%, overall: 68.4%) compared to DW-SPLICE scans (tumor-specific: 26.0%, overall: 56.7%). In terms of image quality, DW-SPLICE is more frequently rated as "good" or "acceptable" (77.0% vs. 56.6%), whereas DW-EPI scans are more commonly rated at the extremes of "poor" (31.9% vs 17.9%) or "excellent" (6.8% vs. 1.2%).

Figure 2. Distribution of rates scores. Frequency distribution of rated scores for (a) tumor distortion (0-3) (n=55, 55 scans), (b) overall distortion (0-3) (n=121, 129 scans) and (c) image quality (1-5) (n=121, 129 scans), as evaluated by the two independent radiologists. Lower scores (0 or 1) denote less distortion for (a) and (b), while higher scores (4 to 5) indicate better image quality for (c).



There was limited but acceptable agreement between raters for the tumor-specific distortion score in DW-EPI and DW-SPLICE, as well as for the overall distortion and image quality score in DW-SPLICE (Table 2). However, for the DW-EPI overall distortion and image quality scores, no agreement was seen when analyzing the weighted Kappa between the raters (Table 2b and 2c) Differences in rater-specific scoring arehighlighted in Supplementary Material Appendix 5.

Within the mixed-effect linear models, which accounted for rater variability and the DWI-pairs, no significant effect of the DWI technique on the tumor-specific distortion score was observed (Table 2a). Yet, there was a significant effect of the DWI technique on the overall distortion score, with DW-SPLICE having an on average lower score, and thus less distortion, compared to DW-EPI (Table 2b). Additionally, DW-SPLICE showed a significantly higher score for image quality, indicating better image quality compared to DW-EPI (Table 2c).

Table 2: Results of the quadratically weighted Kappa and mixed-effect linear models. Scores for (a) tumor specific distortion (0-3), (b) overall distortion (0-3) and (c) image quality (1-5) assessed by two radiologists. The number of MRI scans per subgroup is indicated. Lower scores denote less distortion for (a) and (b), while higher scores indicate better image quality for (c). DW-EPI serves as the reference for DW-SPLICE in the mixed-effects model.

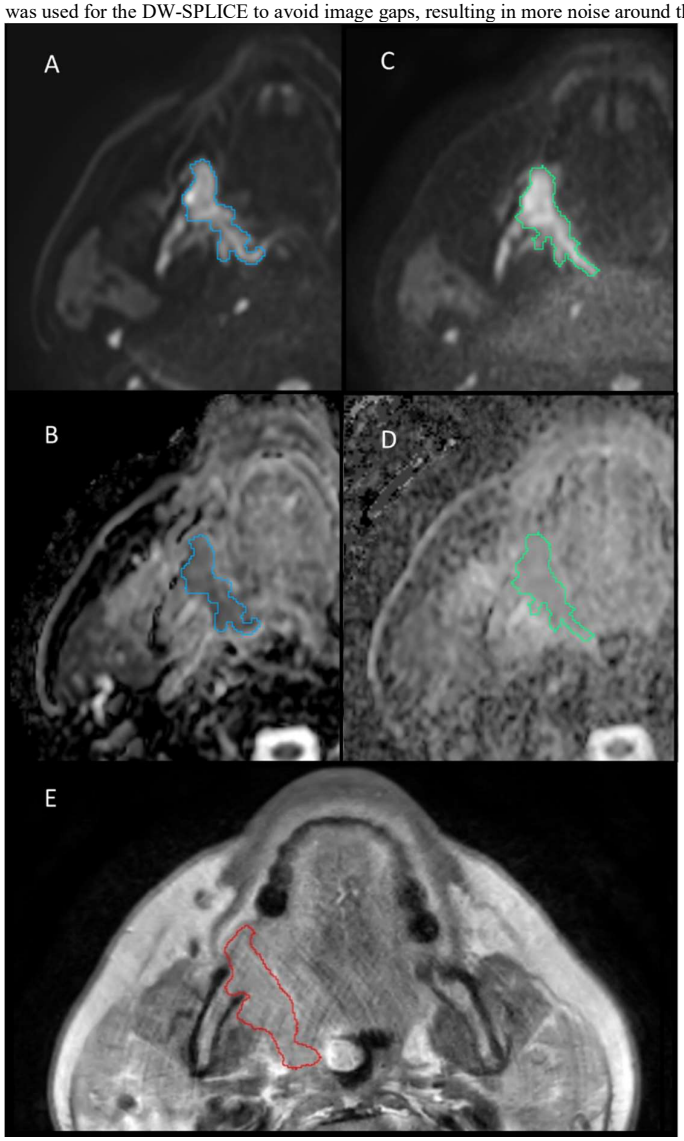
*CI: confidence interval. *significant p-values below the threshold of 0.05*

a. Tumor specific distortion score, n = 55 (55 MRI scan-pairs, 2 reviewers)					
	Kappa		Fixed effect		
	Weighted	*p-value*	*Differential effect (β)*	*95% CI*	*p-value*
DW-EPI	0.62	**<0.001***	Ref		
DW-SPLICE	0.33	**0.004***	-0.163	-0.35-0.02	0.08
b. Overall distortion score, n = 121 (129 MRI scan-pairs, 2 reviewers)					
	Kappa		Fixed effect		
	Weighted	*p-value*	*Differential effect (β)*	*95% CI*	*p-value*
DW-EPI	0.08	0.20	Ref		
DW-SPLICE	0.40	**<0.001***	-0.226	-0.34-0.11	**<0.001***
c. Image quality score, n = 121 (129 MRI scan-pairs, 2 reviewers)					
	Kappa		Fixed effect		
	Weighted	*p-value*	*Differential effect (β)*	*95% CI*	*p-value*
DW-EPI	0.09	0.10	Ref		
DW-SPLICE	0.49	**<0.001***	0.139	0.01-0.27	**0.04***

3.2 Lesion-assessment:

Out of the 55 scans, 36 scans from 36 unique patients showed one or more DWI visible lesion(s) with volume larger than 0.5cm3. Baseline characteristics from this analyzed subgroup and the 56 included lesions (39 benign lesions, 17 malignant lesions) can be found in Supplementary Material Appendix 4. An example for DW-EPI and DW-SPLICE delineation is show in Figure 3.

Figure 3. Comparative delineation on DW-EPI and DW-SPLICE images. Example of DW-EPI delineation shown on (a) b1000 and (b) ADC images, compared with the same case delineated using DW-SPLICE displayed on (c) b1000 and (d) ADC images. The matching T1WIc delineation is shown in (e). DW-SPLICE images exhibit lower contrast compared to DW-EPI, particularly in the ADC map, due to a lower SNR. In contrast, DW-SPLICE offers visually higher geometric similarity to T1WIc compared to the DW-EPI images. Please note that considering the lower SNR, a low threshold for masking was used for the DW-SPLICE to avoid image gaps, resulting in more noise around the image.



The ICC analysis showed a high reliability of the DW-SPLICE and DW-EPI ADC-values across all included lesions and the benign or malignant lesion sub-analyses with agreement and consistency of 0.85 to 0.96(Table 3a and 3b).

Bland-Altman plots and the regression analyses over all lesions and the benign lesion sub-analyses showed a proportional bias in which the difference between DW-EPI and DW-SPLICE ADC increases for higher values (Table 3c, Figure 4A and 4B). This effect was not seen when only the malignant lesions were analyzed (Table 3c, Figure 4C).

Compared to the anatomical T1WIc delineations before registration, the DW-SPLICE delineations demonstrate significantly better geometric similarity with higher DICE scores (0.63 vs 0.47, $p < 0.001$), better recall (0.63 vs 0.45, $p < 0.001$), and greater precision (0.67 vs 0.52, $p < 0.001$) compared to DW-EPI delineations. After registration, these differences dissipate and the scores equalize. See Supplementary Material Appendix 6 for the full results. Figure 3 depicts an example of the T1WIc delineation and both DWI delineations.

Table 3: Intraclass Correlation Coefficient (ICC) results and Bland-Altman analysis of the lesion ADC values. Results of the ICC analyses of the lesion ADC values, assessing (a) agreement and (b) consistency between DW-EPI and DW-SPLICE techniques. And the results of the (c) Bland-

Altman regression analysis results of these lesion ADC values. Results are shown for all lesions, as well as separately for the benign and malignant lesions.
*#: Number of lesions analyzed, ICC: Intraclass Correlation Coefficient, CI: confidence interval. *significant p-values below the threshold of 0.05*

a. Intraclass Correlation Coefficient Agreement

	#	$ICC_{agreement}$	95% CI	p-value
All lesions	56	0.93	0.89-0.96	**<0.001***
Benign	38	0.96	0.92-0.98	**<0.001***
Malignant	18	0.85	0.65-0.94	**<0.001***

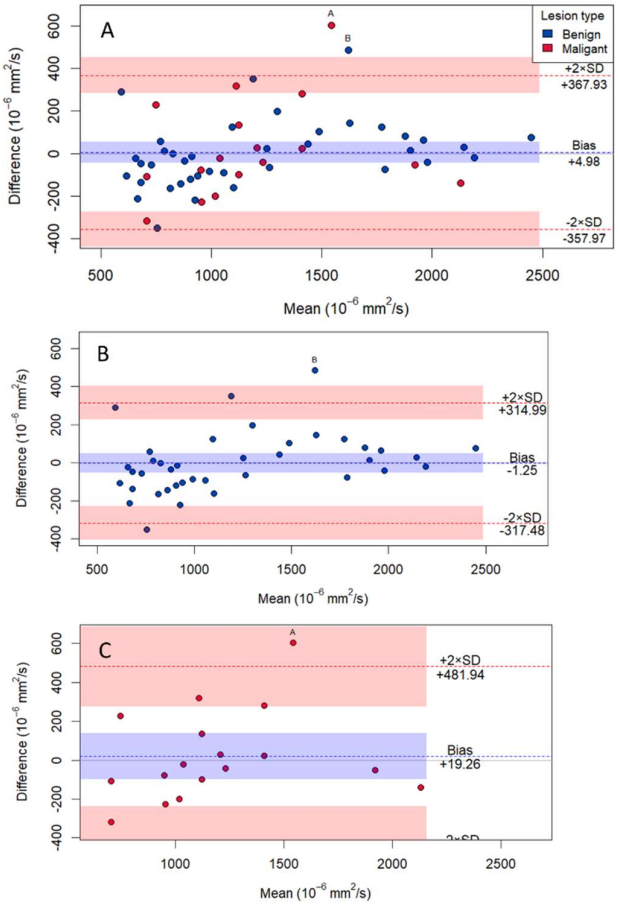
b. Intraclass Correlation Coefficient Consistency

	#	$ICC_{consistency}$	95% CI	p-value
All lesions	56	0.93	0.89-0.96	**<0.001***
Benign	38	0.95	0.91-0.98	**<0.001***
Malignant	18	0.85	0.63-0.94	**<0.001***

c. Bland-Altman regression analyses

	#	Estimate (β)	95% CI	p-value
All lesions	56	0.11	0.008-0.20	**0.03***
Benign	39	0.10	0.009-0.20	**0.03***
Malignant	17	0.34	-0.57-1.26	0.44

Figure 4. Bland-Altman plots of the lesion ADC values. Visual ADC value difference between the DW-EPI (reference) and the DW-SPLICE for (A) all lesions, (B) only the benign lesions and (C) only the malignant lesions.
Label A: a papillary thyroid carcinoma, located low in the neck
Label B: a glomus tumor located low in the neck.



3.3 SCC recurrence assessment:

Of the 66 patients included in the SCC recurrence assessment subgroup, 43.9% had oropharyngeal SCC (OPSCC) and 57.6% received systemic therapy such as radiation and/or chemotherapy. Of the 74 corresponding scans, 52.7% were performed within 6 months after initial treatment. Complete baseline and treatment characteristics can be found in Supplementary Material Appendix 4. There was an overall mean follow-up of 1.5 years post-scan.

DW-SPLICE recurrence scoring showed a better, though not statistically significant, association with actual recurrence at six months (Table 4a) and a statistically significant improvement at one year compared to DW-EPI (Table 4c). A subset analysis focusing on scans with confirmed recurrence further

confirmed the superiority of DW-SPLICE in accurately identifying actual recurrence (Table 4b and d). A significant agreement between the two reviewers scores was seen for DW-EPI and DW-SPLICE (Table 4a).

DW-SPLICE demonstrated a higher sensitivity (47.9% vs. 27.1%) and similar specificity (93.0% vs. 94.0%) in diagnosis recurrence at one year follow-up as compared to DW-EPI. These and other diagnostic accuracy measurements with Receiver Operating Characteristic (ROC) and cross tabular data can be found in Supplementary Material Appendix 7.

Table 4: Cohen's Kappa and mixed linear regression results for (a) recurrence or residual mass at six months and (c) at one year. Sub-analyses of scans displaying recurrence or residual mass are shown for (b) six months and (d) one year.
*CI: confidence interval. *significant p-values below the threshold of 0.05*

a. Full analyses with recurrence up to six months post-scan (scans = 74)					
	Cohen's Kappa	*p-value*	*Odds Ratio*	*95% CI*	*p-value*
DW-EPI	0.48	**<0.001***	Ref		
DW-SPLICE	0.42	**<0.001***	0.72	-0.20-1.65	0.13
b. Subset analyses: Only actual recurrence (scans = 22)					
			Odds Ratio	*95% CI*	*p-value*
		DW-EPI	Ref		
		DW-SPLICE	1.43	0.28-2.79	**0.02***
c. Full analyses with recurrence up to one year post-scan (scans = 74)					
			Odds Ratio	*95% CI*	*p-value*
DW-EPI			Ref		
DW-SPLICE			0.96	0.01-1.92	**0.048***
d. Subset analyses: Only actual recurrence (scans = 24)					
			Odds Ratio	*95% CI*	*p-value*
		DWEPI	Ref		
		DW-SPLICE	1.44	0.34-2.72	**0.02***

DISCUSSION

This study offers a comprehensive evaluation of two DWI techniques (DW-EPI and DW-SPLICE), by examining their differences in measured ADC-values, geometric similarity to T1WIc MRI, and qualitatively scored distortion and image quality. Additionally, we evaluated their effectiveness in detecting recurrences of squamous cell carcinoma (SCC) in the head and neck region. Our findings indicate that DW-SPLICE surpasses DW-EPI on 3T MRI in several areas, including reduced overall image distortion, higher image quality and diagnostic accuracy for detecting recurrent and residual SCC. Thus, integrating DW-SPLICE into standard head and neck MRI protocols can significantly improve tumor assessments.

DW-SPLICE scans showed significantly better image quality and less overall distortion compared to DW-EPI scans (Table 2). Although the tumor-specific distortion score was also better for DW-SPLICE, this difference was not statistically significant, possibly due to the smaller number of patients evaluated for this metric. These findings are in line with the results of Hirata et al.[17] and Fu et al.[19], showing that DW-TSE sequences provide qualitatively better image quality and fewer artifacts than DW-EPI in the oral and ocular regions, respectively.

Our mixed-effect linear models, which accounts for the random effects of the raters, show overall better scores for DW-SPLICE. However, a potential rater bias, as shown by the lack of agreement in DW-EPI's overall distortion and image quality weighted kappa scores shown in Table 2 and further visualized in Supplementary Material Appendix 5, needs to be addressed. Several standardization and training measures such as a kick-off meeting and an evaluation after the first cases were conducted, yet more training may have been necessary or the score should have been simplified. Tumor-specific kappa scores improved after simplifying the scoring into two categories: "no distortion" (including none and mild distortion) and "distortion" (encompassing moderate to severe distortion). Possibly, the time between the standardization measures and the actual evaluations also influenced the results, as disagreements were observed only in the overall distortion and quality scores and less in the tumor-specific distortion score, which was rated only in the first subgroup. (Supplementary Material A4.2) Unlike DW-EPI, DW-SPLICE did not exhibit the same rater disagreement. This discrepancy could stem from the raters' limited familiarity with DW-SPLICE, reducing unconscious biases and, thus, improving adherence to the original scoring guidelines. However, this could also indicate that DW-SPLICE may provide more consistent scores across raters due to better readability and is just less reliant upon experience with the sequence.

As DWI is used as a functional technique, reproducibility of the already established ADC values of DW-EPI is of high importance.[17,20] Within this study, a high degree of ICC reliability is seen between the ADC values of the two DW-techniques. Yet, a proportional bias was observed in ADC measurements of DW-SPLICE compared DW-EPI within the overall and benign lesion analyses.(Table 3 and Figure 4) This indicates that DW-SPLICE ADC values of these groups are higher than the DW-EPI ADC-values, with this difference increasing for higher ADC-values. Previous tests by Schakel et al. that compared ADC values of DW-SPLICE and DW-EPI using ice water phantoms, did corroborate the established literature values.[4] Additionally, a Bland-Altman analyses of Panyarak et al. did not find a proportional bias in salivary glands or lesions in the HN region of DW-TSE compared to DW-EPI.[10] In our study, variations in delineation might have contributed to the observed differences in ADC values. Delineations for DW-EPI and DW-SPLICE were performed independently and blinded, with matching done retrospectively, which could have led to slight inconsistencies. This issue is particularly pronounced in benign lesions, which are harder to delineate due to variations in ADC or b1000 "contrast" and the variety in locations. This is strengthened by the largest discrepancies being observed in lesions located lower in the neck, with less signal from the coil, or that were incompletely depicted (see Figure 4, labels A and B). Moreover, the presence of two lesions visible on DW-EPI but not on DW-SPLICE, and three lesions visible on DW-SPLICE but not on DW-EPI, indicates some variation in lesion visibility between the DWI techniques.

Overall, this suggests that ADC values derived from these two techniques should not be considered interchangeable. Instead, a single technique should be consistently used throughout the staging and follow-up timeline.

While DWI is used as functional imaging technique relying on perceived ADC-value, enhancing geometric accuracy is essential for accurately detecting small lesions, particularly in areas with variable magnetic susceptibility. Reducing susceptibility induced distortion common in DW-EPI and improving geometric similarity helps prevent small lesions from being missed or misidentified as artifacts.[1,4,5] To measure the level of distortion across the two DWI techniques, we compared the overlap between the DWI delineations and the more anatomically accurate T1WIc delineations. As T1WIc and DWI depict different aspects of lesions, a perfect match was not anticipated. However, we postulated that most of the areas discernible on DWI would fall within the T1WIc delineation. In our data, the DW-SPLICE showed a better delineation match compared to DW-EPI before registering the image to the T1WIc (Supplementary Material Appendix 6), corroborating the results of Panyarak et al. which superimposed their DW-TSE and DW-EPI to the T2 weighted imaging (T2W) to calculate a distortion ratio.[10] After registering the DWI to the T1WIc images, this effect dissipates as match percentages equalize, underlining that the same lesion information is likely available in both scans but just hampered by distortion. In clinical radiology practice, where no image registration is applied, radiologists may benefit from using the more accurate DW-SPLICE for locating and assessing lesions.

DWI is beneficial for managing patients with SCC as it is less effected by post-radiotherapy effects and inflammation.[1,20] Therefore, this study focuses on comparing the diagnostic capabilities of DW-EPI and DW-SPLICE in detecting recurrent or residual SCC after treatment. Our results indicate DW-SPLICE was more effective in detecting residual or recurrent SCC at one year follow-up after the assessed MRI scan (p = 0.048). Raters agreed to base their assessments primarily on what was visible on DWI, though the pre-treatment tumor location and stage and an additional sequence (T1WIc) were available for a limited clinical context. Raters lacked access to prior and/or baseline scans and full clinical and treatment details, which are typically considered in standard clinical practice. This shows that DW-SPLICE potentially is a superior technique to assess response.

To optimize DW-image quality we chose only to include scans at 3.0T MRI scanners. Clinical MRI scanners with lower field strengths require longer diffusion gradients, which result in higher TE and noisier DWI. This is especially a problem for DW-TSE like DW-SPLICE, that already has a longer acquisition.[4,17] Considering that clinical MRI scanners are now increasingly equipped with stronger gradient hardware, we opted to test the DW-SPLICE under ideal circumstances.[3] However, as susceptibility artifacts scale with field strength, scanning at higher field strengths may be less advantageous for DW-EPI. This suggests that our results might not be as pronounced at lower field strengths.[17] Furthermore, as SENSE was applied to all DWI scans, standard SNR calculation was not feasible as the distribution of noise is not uniform over the scan.[21]

Other limitations include the retrospective study design resulting in variations in lesion types, time between scan and treatment, and treatment types. In clinical practice, radiographers may adjust the field of view (FOV) during scans. Although the dimensions of the DW-EPI and DW-SPLICE scans are typically linked, in some cases, the FOV of the DW-EPI scan was altered without applying the same adjustment to the corresponding DW-SPLICE scan. This led to more variations in scan duration, resulting in more varied repetition time (TR) for the DW-EPI scans. While these TR variations are not expected to affect the ADC[22], the differences in FOV could potentially have a minor impact on how radiologists interpret the data. A prospective comparison of the two techniques in a more standardized group of patients would be of interest. The reliability of DW-SPLICE for detecting recurrent and residual HNSCC mass at lower field strength should be assessed in future work.

Aside from the DW-SPLICE, several other DWI methods exist to mitigate the artifacts seen for single-shot DW-EPI scans. Either by using a multi-shot EPI sequence or by applying PROPELLER (periodically rotated overlapping parallel lines with enhanced reconstruction) technique for reconstruction.[23,24] Other non-EPI techniques such as HASTE (Half-Fourier acquisition single-shot turbo spin-echo) likewise show value in discerning cholesteatoma, but do not always outperform DW-EPI for HNSCC assessment.[7,25,26] Although DW-SPLICE outperformed the DW-EPI sequence, a comprehensive comparison with all other alternative techniques would be of interest.

CONCLUSIONS

This study shows that DW-SPLICE outperforms DW-EPI in reducing image distortion, enhancing image quality, improving stand-alone diagnostic reliability for detecting recurrent and residual SCC while scanning at high field strength (3.0 Tesla). However, ADC values from these two techniques are not interchangeable; consistent use of a single technique for follow-up is advised. Our data support the integration of DW-SPLICE into clinical practice at 3.0T, as this could improve tumor status assessments. Future work should evaluate the value of DW-SPLICE at lower field strength for full clinical integration.

REFERENCES

1. de Bree, R., *et al*. Detection of locoregional recurrent head and neck cancer after (chemo)radiotherapy using modern imaging. *Oral Oncology* **45**, 386-393 (2009).
2. Driessen, J.P., *et al*. Diffusion-weighted imaging in head and neck squamous cell carcinomas: A systematic review. *Head & Neck* **37**, 440-448 (2015).
3. Iima, M., Partridge, S.C. & Le Bihan, D. Six DWI questions you always wanted to know but were afraid to ask: clinical relevance for breast diffusion MRI. *Eur Radiol* **30**, 2561-2570 (2020).
4. Schakel, T., Hoogduin, J.M., Terhaard, C.H.J. & Philippens, M.E.P. Technical Note: Diffusion-weighted MRI with minimal distortion in head-and-neck radiotherapy using a turbo spin echo acquisition method. *Med Phys* **44**, 4188-4193 (2017).
5. Schakel, T., Hoogduin, J.M., Terhaard, C.H. & Philippens, M.E. Diffusion weighted MRI in head-and-neck cancer: geometrical accuracy. *Radiother Oncol* **109**, 394-397 (2013).
6. van Egmond, S.L., Stegeman, I., Grolman, W. & Aarts, M.C. A Systematic Review of Non-Echo Planar Diffusion-Weighted Magnetic Resonance Imaging for Detection of Primary and Postoperative Cholesteatoma. *Otolaryngol*

Head Neck Surg **154**, 233-240 (2016).

7. Verhappen, M.H., *et al*. Diffusion-weighted MR imaging in head and neck cancer: comparison between half-fourier acquired single-shot turbo spin-echo and EPI techniques. *AJNR Am J Neuroradiol* **33**, 1239-1246 (2012).
8. Sakamoto, J., *et al*. Comparison of accuracy of intravoxel incoherent motion and apparent diffusion coefficient techniques for predicting malignancy of head and neck tumors using half-Fourier single-shot turbo spin-echo diffusion-weighted imaging. *Magn Reson Imaging* **32**, 860-866 (2014).
9. Sakamoto, J., Sasaki, Y., Otonari-Yamamoto, M. & Sano, T. Comparison of various methods for quantification of apparent diffusion coefficient of head and neck lesions with HASTE diffusion-weighted MR imaging. *Oral Surg Oral Med Oral Pathol Oral Radiol* **114**, 266-276 (2012).
10. Panyarak, W., Chikui, T., Yamashita, Y., Kamitani, T. & Yoshiura, K. Image Quality and ADC Assessment in Turbo Spin-Echo and Echo-Planar Diffusion-Weighted MR Imaging of Tumors of the Head and Neck. *Acad Radiol* **26**, e305-e316 (2019).
11. McDonald, B.A., *et al*. In Vivo and Phantom Repeatability of Diffusion-Weighted MRI Sequences on 1.5T MRI-Linear Accelerator (MR-Linac) and MR Simulator Devices for Head and Neck Cancers: Results from a Prospective R-IDEAL Stage 2a Evaluation of Tumor and Normal Tissue Apparent Diffusion Coefficients as Quantitative Imaging Biomarkers. *medRxiv*, 2022.2005.2028.22275724 (2022).
12. Sakamoto, J., *et al*. Tissue characterization of head and neck lesions using diffusion-weighted MR imaging with SPLICE. *Eur J Radiol* **69**, 260-268 (2009).
13. Schick, F. SPLICE: sub-second diffusion-sensitive MR imaging using a modified fast spin-echo acquisition mode. *Magn Reson Med* **38**, 638-644 (1997).
14. Deng, J., Omary, R.A. & Larson, A.C. Multishot diffusion-weighted SPLICE PROPELLER MRI of the abdomen. *Magnetic Resonance in Medicine* **59**, 947-953 (2008).
15. Cohen, J.F., *et al*. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* **6**, e012799 (2016).
16. Wang, J., *et al*. Head and neck lesions: characterization with diffusion-weighted echo-planar MR imaging. *Radiology* **220**, 621-630 (2001).
17. Hirata, K., *et al*. Comparison of the image quality of turbo spin echo- and echo-planar diffusion-weighted images of the oral cavity. *Medicine (Baltimore)* **97**, e0447 (2018).
18. Kojima, T., *et al*. Efficacy of the radial acquisition regime (RADAR) for acquiring head and neck MR images. *Br J Radiol* **89**, 20160007 (2016).
19. Fu, Q., *et al*. Turbo Gradient and Spin-Echo BLADE-DWI for Extraocular Muscles in Thyroid-Associated Ophthalmopathy. *J Clin Med* **12**(2023).
20. Thoeny, H.C., De Keyzer, F. & King, A.D. Diffusion-weighted MR imaging in the head and neck. *Radiology* **263**, 19-32 (2012).
21. Dietrich, O., Raya, J.G., Reeder, S.B., Reiser, M.F. & Schoenberg, S.O. Measurement of signal-to-noise ratios in MR images: influence of multichannel coils, parallel imaging, and reconstruction filters. *J Magn Reson Imaging* **26**, 375-385 (2007).
22. Celik, A. Effect of imaging parameters on the accuracy of apparent diffusion coefficient and optimization strategies. *Diagn Interv Radiol* **22**, 101-107 (2016).
23. Wang, F.N., *et al*. PROPELLER EPI: an MRI technique suitable for diffusion tensor imaging at high field strength with reduced geometric distortions. *Magn Reson Med* **54**, 1232-1240 (2005).
24. Mavroidis, P., *et al*. Comparison Between EPI DWI and PROPELLER DWI in Brain MR Imaging. *Current Problems in Diagnostic Radiology* **53**, 73-80 (2024).
25. Schouten, C.S., *et al*. Diffusion-weighted EPI- and HASTE-MRI and 18F-FDG-PET-CT early during chemoradiotherapy in advanced head and neck cancer. *Quantitative Imaging in Medicine and Surgery* **4**, 239-250 (2014).
26. Benson, J.C., Carlson, M.L. & Lane, J.I. Non-EPI versus Multishot EPI DWI in Cholesteatoma Detection: Correlation with Operative Findings. *AJNR Am J Neuroradiol* **42**, 573-577 (2021).

SUPPLEMENTAL FILES

Table of content:

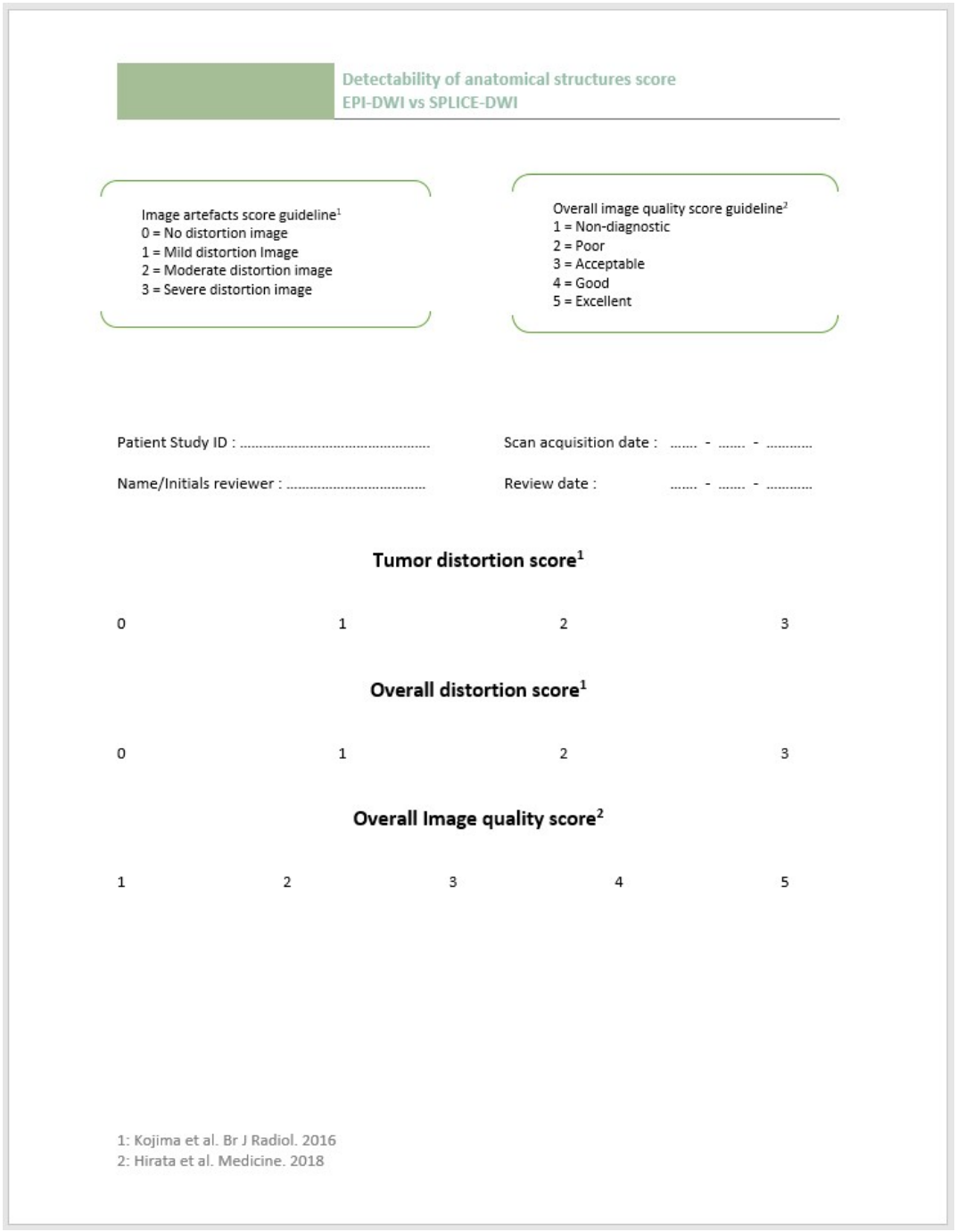
APPENDIX 1 – STANDARDS FOR REPORTING DIAGNOSTIC ACCURACY STUDIES (STARD) 2015 CHECKLIST AND APPLICABILITY TO THIS MANUSCRIPT.

Section & Topic	NO	Item	Applied	Elaboration, if needed
TITLE OR ABSTRACT				
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy	Yes	Measures of diagnostic value comparison mentioned in abstract: mixed-effect linear models, mixed logistic regression inter-class correlation.
ABSTRACT				
	2	Structured summary of study design, methods, results, and conclusions	Yes	-
INTRODUCTION				
	3	Scientific and clinical background, including the intended use and clinical role of the index test	Yes	-
	4	Study objectives and hypotheses	Yes	-
METHODS				
Study design	5	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	Yes	The retrospective design is indicated in the materials and methods; 2.1 Population and subgroups.
Participants	6	Eligibility criteria	Yes	Eligibility criteria are indicated in the materials and methods; 2.1 Population and subgroups.
	7	On what basis potentially eligible participants were identified	Yes	Basis for inclusion is mentioned in the materials and methods; 2.1 Population and subgroups.
	8	Where and when potentially eligible participants were identified	Yes	Inclusion time and location is noted in the materials and methods; 2.1 Population and subgroups.
	9	Whether participants formed a consecutive, random or convenience series	Yes	As indicated in the materials and methods; 2.1 Population and subgroups, a consecutive inclusion was conducted.
Test methods	10a	Index test, in sufficient detail to allow replication	Yes	Details of the scan highlighted in Table 1
	10b	Reference standard, in sufficient detail to allow replication	Yes	Details of the scan highlighted in Table 1
	11	Rationale for choosing the reference standard	Yes	The current standard in clinical practice is chosen a reference standard. Elaborated in the Introduction.
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	Yes	Qualitative groups as defined in previous research. See Material and Methods; 2.5 Qualitative evaluation. No other cut-off values applied
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	Yes	Equal to 12a
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	Yes	Defined in material and methods; 2.3 Data preparation, 2.4 Delineation and registration and in the discussion
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	Yes	Same as 13a
Analysis	14	Methods for estimating or comparing measures of diagnostic accuracy	Yes	Defined in materials and methods; 2.6 Statistical analyses
	15	How indeterminate index test or reference standard results were handled	Yes	Because of the pairwise analyses, lesions for the lesions assessment subgroup were only included if they wer visible on either the b1000 or ADC of both the DW-EPI and DW-SPLICE scan. Indicated in materials and methods; 2.1 Population and subgroups. No other indetermined test results are expected
	16	How missing data on the index test and reference standard were handled	Yes	Only limited follow-up data was available for some of the patients in the recurrence assessment subgroup, as this is equal for the DW-EPI and DW-SPLCIE, no further correction is applied. See Supplementary Material Appendix 4
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	Yes	The research was conducted based on a pre-specified design. Yet some larger variations in scan parameters are noted (Table 1).
	18	Intended sample size and how it was determined	No	Not defined in the manuscript.

RESULTS				
Participants	19	Flow of participants, using a diagram	Yes	See Figure 1
	20	Baseline demographic and clinical characteristics of participants	Yes	See Supplementary Material Appendix 4
	21a	Distribution of severity of disease in those with the target condition	Yes	Any stage of HNSCC was included for the recurrence assessment.
	21b	Distribution of alternative diagnoses in those without the target condition	Yes	For the lesion assessment, both malignant and Benign lesions were included. For the recurrence assessment, patients with or without recurrence were included.
	22	Time interval and any clinical interventions between index test and reference standard	-	Acquired at the same MRI
Test results	23	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	Yes	See Supplementary Material Appendix 7
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	Yes	Estimates and precision are reported in Table 2, 3, and 4 as well as in Supplementary Material Appendix 6 and 7
	25	Any adverse events from performing the index test or the reference standard	-	None to be expected
DISCUSSION				
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalizability	Yes	-
	27	Implications for practice, including the intended use and clinical role of the index test	Yes	-
OTHER INFORMATION				
	28	Registration number and name of registry	-	Not performed as clinical trial. Registered as IRB under IRBd22-270.
	29	Where the full study protocol can be accessed	No	No other protocol available then listed in the manuscript
	30	Sources of funding and other support; role of funders	Yes	No disclosures of funding or other support

APPENDIX 2 – IMAGE QUALITY, DISTORTION AND RECURRENCE SCORE FORMS

A2.1 Lesion assessment subgroup score forms
An excel adaptation of this score form was used to record the answers.



Detectability of anatomical structures score
EPI-DWI vs SPLICE-DWI

Image artefacts score guideline[1]
0 = No distortion image
1 = Mild distortion Image
2 = Moderate distortion image
3 = Severe distortion image

Overall image quality score guideline[2]
1 = Non-diagnostic
2 = Poor
3 = Acceptable
4 = Good
5 = Excellent

Patient Study ID : .....

Scan acquisition date : ..... - ..... - .....

Name/Initials reviewer : .....

Review date : ..... - ..... - .....

Tumor distortion score[1]

0 1 2 3

Overall distortion score[1]

0 1 2 3

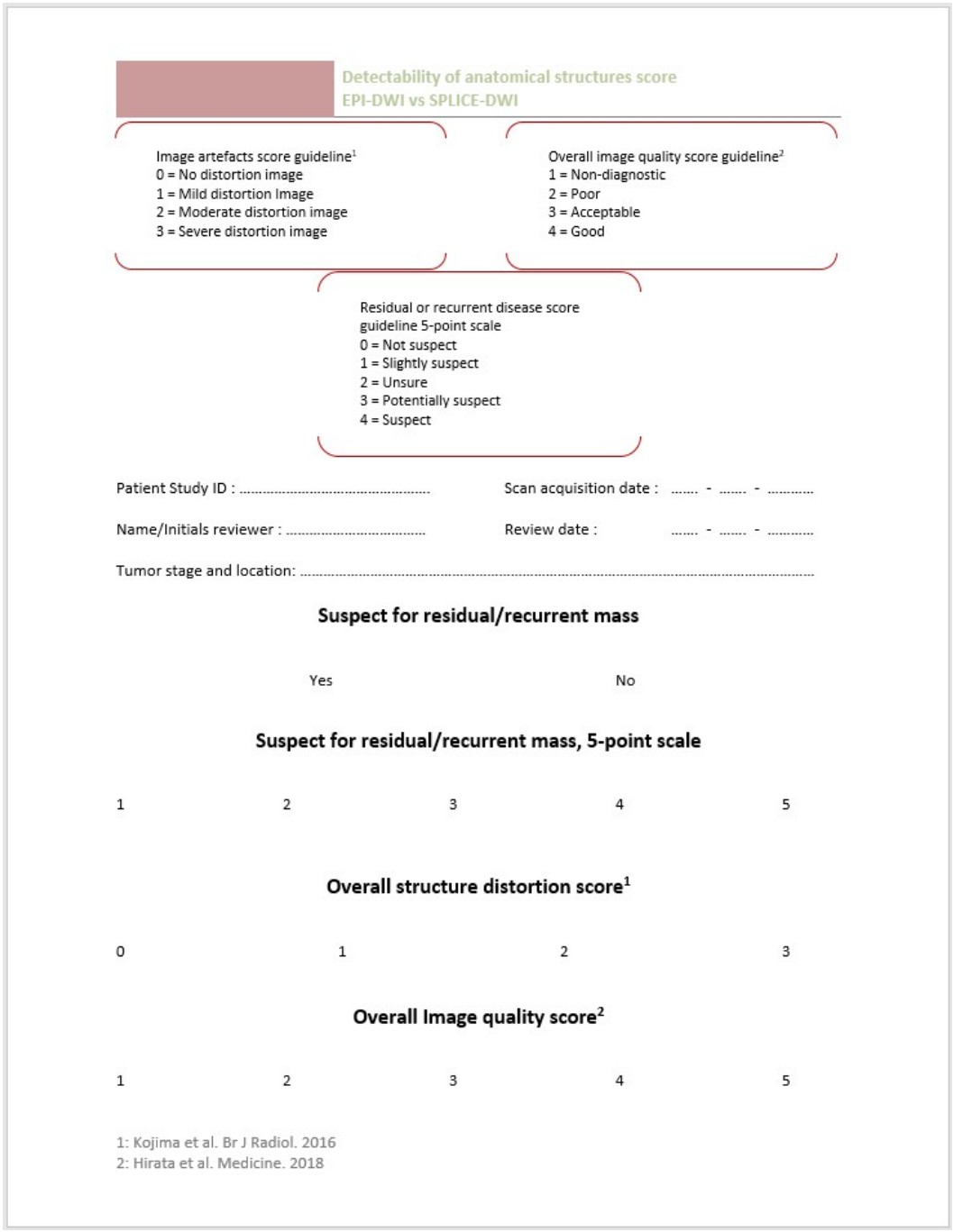
Overall Image quality score[2]

1 2 3 4 5

1: Kojima et al. Br J Radiol. 2016
2: Hirata et al. Medicine. 2018

A2.2 SCC recurrence assessment subgroup score forms

An excel adaptation of this score form was used to record the answers. The tumor location and stage were provided to the reviewers.



Detectability of anatomical structures score
EPI-DWI vs SPLICE-DWI

Image artefacts score guideline[1]
0 = No distortion image
1 = Mild distortion Image
2 = Moderate distortion image
3 = Severe distortion image

Overall image quality score guideline[2]
1 = Non-diagnostic
2 = Poor
3 = Acceptable
4 = Good

Residual or recurrent disease score
guideline 5-point scale
0 = Not suspect
1 = Slightly suspect
2 = Unsure
3 = Potentially suspect
4 = Suspect

Patient Study ID : ..... Scan acquisition date : ..... - ..... - .....

Name/Initials reviewer : ..... Review date : ..... - ..... - .....

Tumor stage and location: .....

Suspect for residual/recurrent mass

Yes No

Suspect for residual/recurrent mass, 5-point scale

1 2 3 4 5

Overall structure distortion score[1]

0 1 2 3

Overall Image quality score[2]

1 2 3 4 5

1: Kojima et al. Br J Radiol. 2016
2: Hirata et al. Medicine. 2018

APPENDIX 3 – FORMULAS

1. **DICE Similarity Coefficient**

$$\boldsymbol{DSC} = \frac{\mathbf{2}|X \cap Y|}{|\boldsymbol{X}|+|\boldsymbol{Y}|} \qquad \textbf{Formula (1)}$$

DSC= DICE Similarity Coefficient; X= DWI delineation; Y = T1Wc delineation; $|X \cap Y|$= Overlap between DWI and T1Wc delineation;

2. **Recall score**

$$\boldsymbol{Recall\ score} = \frac{|X \cap Y|}{|\boldsymbol{Y}|} \qquad \textbf{Formula (2)}$$

X= DWI delineation; Y = T1Wc delineation; $|X \cap Y|$= Overlap between DWI and T1Wc delineation;

3. **Precision score**

$$\boldsymbol{Precision\ score} = \frac{|X \cap Y|}{|\boldsymbol{X}|} \qquad \textbf{Formula (3)}$$

X= DWI delineation; Y = T1Wc delineation; $|X \cap Y|$= Overlap between DWI and T1Wc delineation;

APPENDIX 4 - PATIENT CHARACTERISTICS OF ALL SUBGROUPS

A4.1 Patient characteristics for (A) the overall analyses, (B) the lesion-assessment subgroup and (C) the SCC recurrence subgroup.

*SCC: Squamous cell carcinoma, NOS: not otherwise specified, N: Total patients in cohort, n: number of patients in subgroup, *Of the 55 patients in the lesion assessment subgroup, 36 had lesions >0.5cm3 visible on DWI. †Patients can have both benign and malignant lesions. ‡the patient is already being treated for a recurrent lesion at the time of the scan.*

A. Overall baseline characteristics		(N = 121)
Age at time scan (years)	Mean (SD)	62.2 ± 13.5
Sex	Male	82 (67.8%)
	Female	39 (32.2%)

B. Lesion-assessment subgroup baseline characteristics		(n = 36)*
Age at time scan (years)	Mean (SD)	57.7 ± 15.0
Sex	Male	17 (47.2%)
	Female	19 (52.8%)
Total lesions	DWI-visible, >0,5cm^3	56 (n=36)
Lesion classification	Benign	39 (n=23†)
	Malignant	17 (n=13†)
Lesions: Malignant	HNSCC	4 (23.5%)
	Lymph node metastasis	4 (23.5%)
	Papillary thyroid carcinoma	3 (17.7%)
	Sarcoma	2 (11.8%)
	Basal cell carcinoma	2 (11.8%)
	Cutaneous SCC	1 (5.9%)
	Mucoepidermoid carcinoma	1 (5.9%)
Lesions: Benign	Salivary gland tumors	13 (33.3%)
	Enlarged but (PA) normal lymph node	10 (25.6%)
	Cyst NOS	5 (12.8%)
	Reactive tissue	4 (10.3%)
	Benign neoplasm NOS	2 (5.1%)
	Glomus tumor	3 (7.7%)
	Schwannoma	1 (2.6%)
	IgG4-related disease	1 (2.6%)

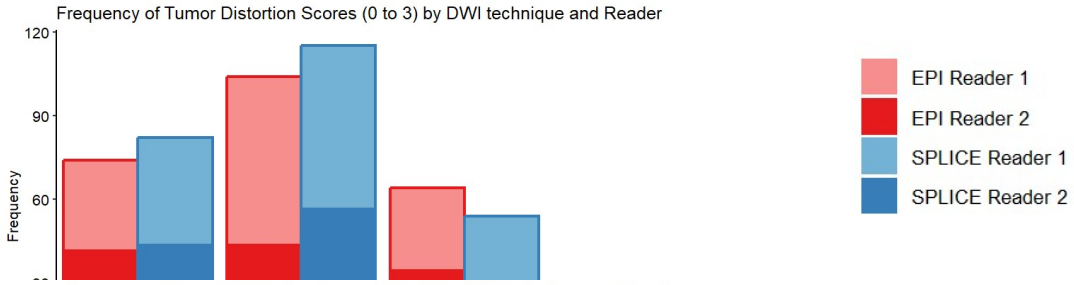
C. SCC recurrence subgroup baseline and treatment characteristics		(n = 66)
Age at time scan (years)	Mean (SD)	65.5 ± 10.4
Sex	Male	53 (80.3%)
	Female	13 (19.7%)
Tumor location	Oral cavity	13 (19.7%)
	Oropharynx	29 (43.9%)
	Hypopharynx	3 (4.5%)
	Larynx	5 (7.6%)
	Nasopharynx	5 (7.6%)
	Cutaneous	7 (10.6%)
	Unknown Primary SCC	4 (6.1%)
Pre-treatment Clinical T-stage	T1	15 (22.7%)
	T2	13 (19.7%)
	T3	10 (15.2%)
	T4	11 (16.7%)
	Recurrent‡	13 (19.7%)
	Unknown Primary	4 (6.1%)
Pre-treatment Clinical N-stage	N0	16(24.2%)
	N1	14 (21.2%)
	N2	16 (24.2%)
	N3	7 (10.6%)
	Recurrent‡	13 (19.7%)
Treatment regiment	Surgery	15 (22.7%)
	Systemic therapy: Radiation/chemotherapy	38 (57.6%)
	Surgery + Systemic therapy	7 (10.6%)
	Immunotherapy + Surgery and/or systemic	6 (9.1%)
Time between first treatment and MRI	0-6 months	39 (n = 38)
	7-12 months	10 (n = 9)
	12+ months	25 (n = 19)

APPENDIX 5 – READER SPECIFIC SCORING

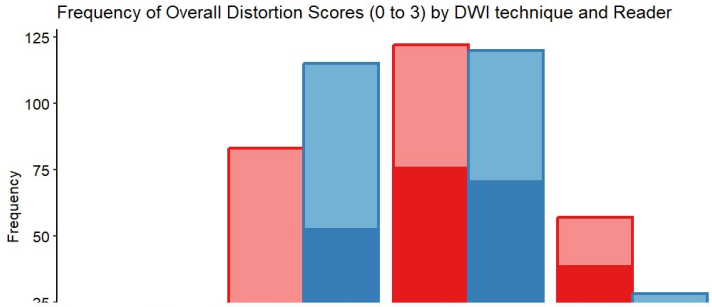
A5.1. Frequencies of rated scores split by rater

Frequencies of rated scores for (a) tumor distortion (n=55, 55 scans), (b) overall distortion (n=121, 129 scans) and (c) image quality (n=121, 129 scans) split by the two independent readers.

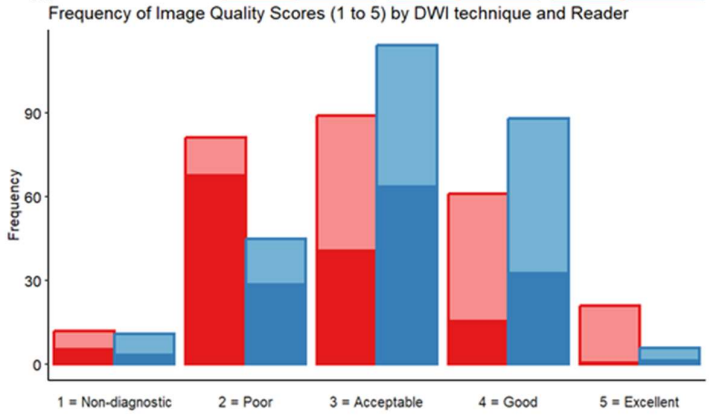
a.



b.



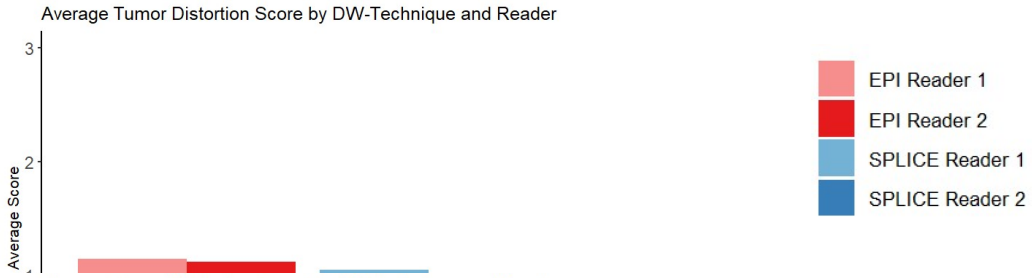
c.



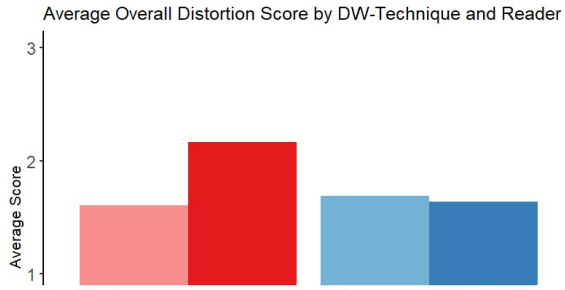
A5.2. Average of rated scores split by rater

Average of rated scores for (a) tumor distortion (n=55, 55 scans), (b) overall distortion (n=121, 129 scans) and (c) image quality (n=121, 129 scans) split by the two independent readers and sorted by DW-technique.

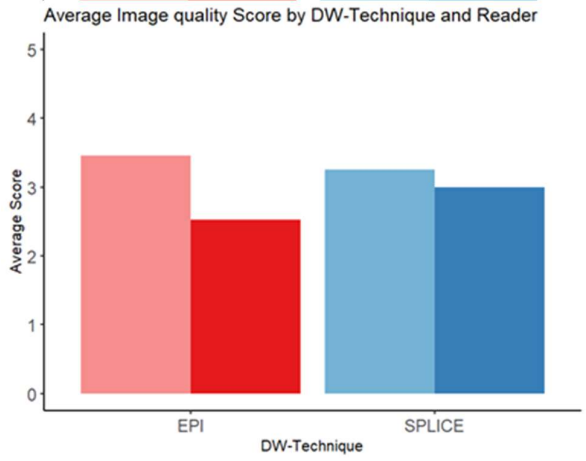
A.



B.



C.



APPENDIX 6 – DICE SCORE, RECALL SCORE AND PRECISION SCORE BEFORE AND AFTER REGISTRATION

A6.1. Results for DICE, recall, and precision scores are listed for DW-EPI and DW-SPLICE delineations (a) before and (b) after registration to the T1Wc. *SD: Standard deviation.*

a. Before registration of DWI to T1Wc

	DICE	*SD*	*p-value*	*Recall*	*SD*	*p-value*	*Precision*	*SD*	*p-value*
DW-SPLICE	0.63	0.16		0.63	0.16		0.67	0.20	
DW-EPI	0.47	0.21		0.45	0.22		0.52	0.24	
			< 0.001			**< 0.001**			**< 0.001**

b. After registration of DWI to T1Wc

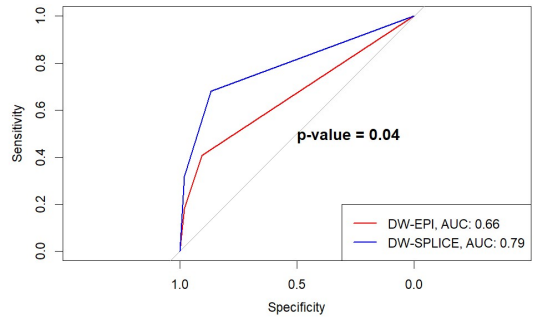
	DICE	*SD*	*p-value*	*Recall*	*SD*	*p-value*	*Precision*	*SD*	*p-value*
DW-SPLICE	0.48	0.18		0.46	0.20		0.57	0.23	
DW-EPI	0.44	0.21		0.42	0.22		0.52	0.27	
			0.06			0.11			0.10

APPENDIX 7 – DIAGNOSTIC ACCURACY MEASURES FOR RECURRENCE ASSESSMENT

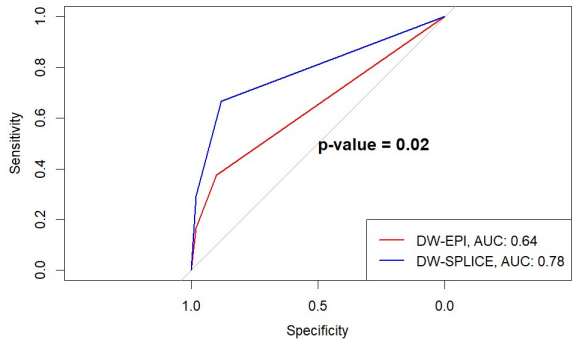
A7.1. ROC-curve of recurrence measurement by DW-EPI and DW-SPLICE.

Each DW-technique is assessed by comparing the reader-average recurrence score to actual recurrences within (a) six months post-scan or (b) one year post-scan, using Receiver Operating Characteristic (ROC) curves. The area under the curve (AUC) quantifies the diagnostic accuracy. Differences between DW techniques are statistically compared using the DeLong test, *p*-value is shown.

a.



b.



A7.2. Cross tables and performance characteristics

Tables presenting cross-tabulated data and diagnostic metrics—sensitivity, specificity, negative predictive value, and positive predictive value—for (a, c) DW-EPI and (b, c) DW-SPLICE techniques, comparing predicted to actual recurrences within six months post-scan (a, b) and one year post-scan (c, d). All 74 scans were evaluated by two independent reviewers, totalling 148 ratings.

a. DW-EPI, recurrences within six months post-scan

		Actual	
		No recurrence	Recurrence
Predicted	No recurrence	98	31
	Recurrence	6	13

Sensitivity	29.5%
Specificity	94.2%
Positive Predictive Value	68.4%
Negative Predictive Value	76.0%

b. DW-SPLICE, recurrences within six months post-scan

		Actual	
		No recurrence	Recurrence
Predicted	No recurrence	96	22
	Recurrence	8	22

Sensitivity	50.0%
Specificity	92.3%
Positive Predictive Value	73.3%
Negative Predictive Value	81.4%

c. DW-EPI, recurrences within one year post-scan

		Actual	
		No recurrence	Recurrence
Predicted	No recurrence	94	35
	Recurrence	6	13

Sensitivity	27.1%
Specificity	94.0%
Positive Predictive Value	68.4%
Negative Predictive Value	72.9%

d. DW-SPLICE, recurrences within one year post-scan

		Actual	
		No recurrence	Recurrence
Predicted	No recurrence	93	25
	Recurrence	7	23

Sensitivity	47.9%
Specificity	93.0%
Positive Predictive Value	76.7%
Negative Predictive Value	78.8%