

Are your MRI contrast agents cost-effective?

Learn more about generic Gadolinium-Based Contrast Agents.



FRESENIUS
KABI

caring for life

AJNR

MRI with DWI for the Detection of Posttreatment Head and Neck Squamous Cell Carcinoma: Why Morphologic MRI Criteria Matter

A. Ailianou, P. Mundada, T. De Perrot, M. Puszczazieri, P.-A. Poletti and M. Becker

This information is current as of April 19, 2024.

AJNR Am J Neuroradiol 2018, 39 (4) 748-755

doi: <https://doi.org/10.3174/ajnr.A5548>

<http://www.ajnr.org/content/39/4/748>

MRI with DWI for the Detection of Posttreatment Head and Neck Squamous Cell Carcinoma: Why Morphologic MRI Criteria Matter

A. Ailianou, P. Mundada, T. De Perrot, M. Pusztaszieri, P.-A. Poletti, and M. Becker



ABSTRACT

BACKGROUND AND PURPOSE: Although diffusion-weighted imaging combined with morphologic MRI (DWIMRI) is used to detect posttreatment recurrent and second primary head and neck squamous cell carcinoma, the diagnostic criteria used so far have not been clarified. We hypothesized that precise MRI criteria based on signal intensity patterns on T2 and contrast-enhanced T1 complement DWI and therefore improve the diagnostic performance of DWIMRI.

MATERIALS AND METHODS: We analyzed 1.5T MRI examinations of 100 consecutive patients treated with radiation therapy with or without additional surgery for head and neck squamous cell carcinoma. MRI examinations included morphologic sequences and DWI ($b=0$ and $b=1000$ s/mm²). Histology and follow-up served as the standard of reference. Two experienced readers, blinded to clinical/histologic/follow-up data, evaluated images according to clearly defined criteria for the diagnosis of recurrent head and neck squamous cell carcinoma/second primary head and neck squamous cell carcinoma occurring after treatment, post-radiation therapy inflammatory edema, and late fibrosis. DWI analysis included qualitative (visual) and quantitative evaluation with an ADC threshold.

RESULTS: Recurrent head and neck squamous cell carcinoma/second primary head and neck squamous cell carcinoma occurring after treatment was present in 36 patients, whereas 64 patients had post-radiation therapy lesions only. The Cohen κ for differentiating tumor from post-radiation therapy lesions with MRI and qualitative DWIMRI was 0.822 and 0.881, respectively. Mean ADC_{mean} in recurrent head and neck squamous cell carcinoma/second primary head and neck squamous cell carcinoma occurring after treatment ($1.097 \pm 0.295 \times 10^{-3}$ mm²/s) was significantly lower ($P < .05$) than in post-radiation therapy inflammatory edema ($1.754 \pm 0.343 \times 10^{-3}$ mm²/s); however, it was similar to that in late fibrosis ($0.987 \pm 0.264 \times 10^{-3}$ mm²/s, $P > .05$). Although ADCs were similar in tumors and late fibrosis, morphologic MRI criteria facilitated distinction between the 2 conditions. The sensitivity, specificity, positive and negative predictive values, and positive and negative likelihood ratios (95% CI) of DWIMRI with ADC_{mean} $< 1.22 \times 10^{-3}$ mm²/s and precise MRI criteria were 92.1% (83.5–100.0), 95.4% (90.3–100.0), 92.1% (83.5–100.0), 95.4% (90.2–100.0), 19.9 (6.58–60.5), and 0.08 (0.03–0.24), respectively, indicating a good diagnostic performance to rule in and rule out disease.

CONCLUSIONS: Adding precise morphologic MRI criteria to quantitative DWI enables reproducible and accurate detection of recurrent head and neck squamous cell carcinoma/second primary head and neck squamous cell carcinoma occurring after treatment.

ABBREVIATIONS: DWIMRI = combined MRI with morphologic sequences and DWI; HN = head and neck; HNSCC = head and neck squamous cell carcinoma; LR = likelihood ratio; pHNSCC = primary head and neck squamous cell carcinoma; rHNSCC = recurrent head and neck squamous cell carcinoma; RTH = radiation therapy; sHNSCC = second primary head and neck squamous cell carcinoma occurring after treatment


Squamous cell carcinoma accounts for 95% of all cancers in the head and neck (HN) region.¹ The aim of treatment is to provide a cure with the maximum possible preservation of organ function. Treatment options comprise radiation therapy (RTH) and/or chemotherapy and/or an operation.^{1,2} Up to 25% of early-


stage primary head and neck squamous cell carcinomas (pHNSCCs) and up to 50% of advanced-stage tumors show treatment failure in the form of residual/recurrent HNSCC (rHNSCC), whereas 15% of patients develop second primary HN cancer (sHNSCC).^{2,3} Early

Received July 28, 2017; accepted after revision December 2.

From the Division of Radiology (A.A., P.M., T.D.P., P.-A.P., M.B.), Department of Imaging and Medical Informatics, and Division of Clinical Pathology (M.P.), Department of Laboratory and Genetics, Geneva University Hospitals, University of Geneva, Geneva, Switzerland. Dr Ailianou is currently affiliated with the Hospital of Neuchâtel, Neuchâtel, Switzerland.

Please address correspondence to Minerva Becker, MD, Division of Radiology, Department of Imaging and Medical Informatics, Geneva University Hospitals, Rue Gabrielle-Perret-Gentil 4, 1211 Geneva 14, Switzerland; e-mail: Minerva.Becker@hcuge.ch

 Indicates article with supplemental on-line table.

 Indicates article with supplemental on-line photos.

<http://dx.doi.org/10.3174/ajnr.A5548>

detection of rHNSCC/sHNSCC prompts curative salvage treatment and may allow preservation of organ function.

RTH causes dramatic gross and microenvironmental changes in the affected soft tissues. In most cases, inflammatory edema develops, which leads to formation of fibrosis and scar tissue. Inflammatory edema and scar tissue may coexist, and they may render clinical/endoscopic evaluation challenging. Clinical/endoscopic examinations may miss a tumor under intact mucosa or may underestimate it due to its multifocal recurrence pattern. Moreover, panendoscopy is an invasive procedure requiring general anesthesia. A biopsy with negative findings does not exclude rHNSCC/sHNSCC, and multiple biopsies may increase overall morbidity.^{2,4,5}

MRI with morphologic sequences and diffusion-weighted imaging (DWIMRI) provides anatomic, qualitative, and quantitative functional information allowing noninvasive assessment of treatment response and detection of rHNSCC/sHNSCC.⁵⁻⁹ The imaging features of rHNSCC on morphologic MRI sequences have been scarcely addressed in the literature,^{1,5,7,10} and several studies have focused on demonstrating the utility of DWI for the detection of recurrence.^{5-9,11-13} However, the diagnostic performance of clearly defined morphologic MRI criteria, the interobserver reproducibility, and the impact of each technique (morphologic MRI versus combined DWI with morphologic MRI) have not been assessed in these studies. Earlier studies were rather small,^{6,7,13} and they were based on a mix of HNSCC and other histologic tumor types^{7,11} or included both local and nodal recurrence.^{6,12}

The aim of this study was to evaluate the diagnostic performance of morphologic MRI with clearly defined criteria for the detection of rHNSCC/sHNSCC and to compare it with the diagnostic performance of DWI alone and of DWI combined with morphologic MRI. We likewise aimed to analyze the added value of each technique and their respective contributions to the overall diagnostic performance of DWIMRI.

MATERIALS AND METHODS

Patients

The institutional ethics committee approved this retrospective study, which was performed in accordance with the guidelines of the Declaration of Helsinki II. Informed consent was waived. The study population included 114 consecutive patients seen during 3.5 years in our institution. The patients underwent 1.5T MRI with morphologic MRI and DWI sequences for clinically suspected recurrence or for follow-up after RTH with or without an additional operation. Fourteen patients were excluded because of absent follow-up ($n = 6$), nondiagnostic DWI ($n = 5$), and nondiagnostic morphologic MRI and DWI ($n = 3$). The remaining 100 patients formed the basis of the current study. The series comprised 21 women and 79 men with a median age (25 quartile–75 quartile) of 62 years (55–69 years), ranging from 18 to 89 years. The median time interval between the end of RTH and MRI was 14 months (4.5–51 months).

Imaging Technique

The MRI protocol included imaging from the skull base to the thoracic inlet with FSE T2 and FSE T1 sequences in the axial plane, a short τ inversion recovery sequence in the coronal plane, and axial spin-echo EPI DWI ($b=0$ and 1000 s/mm^2) with apparent

diffusion coefficient map calculation using monoexponential fitting. After administration of a gadolinium-based contrast agent, we performed T1WI \pm fat saturation in the axial, coronal, and/or sagittal plane. Sequence parameters were the following: TR/TE = 3300/106 ms, 24 slices, slice thickness = 3 mm, FOV = 230×180 mm, matrix = 512×416 , acquisition time = 3 minutes 30 seconds; and TR/TE = 771/11 ms, 30 slices, slice thickness = 3–4 mm, FOV = 230×230 mm, matrix = 512×512 , acquisition time = 3 minutes 56 seconds for T2- and T1-weighted acquisitions, respectively. DWI parameters were the following: TR/TE = 3200/86 ms, 40 slices, slice thickness = 3–4 mm, FOV = 230×230 mm, matrix = 128×128 , acquisition time = 3 minutes 2 seconds.

Evaluation of Morphologic MRI and DWI and Applied Diagnostic Criteria

Two board-certified HN radiologists (>10 years' experience in HN oncology), blinded to all clinical data (patient history, histology, and clinical/radiologic follow-up), analyzed the images separately. First, all morphologic MRIs were evaluated with the reader blinded to DWI. Then the readers evaluated morphologic MRI and DWI together (DWIMRI). After separate evaluations by each reader and assessment of interobserver variability, consensus was reached in cases of discrepant interpretations.

On the basis of the experience at our institution and according to published data,^{1,14,15} the interpretation of morphologic MRI was performed according to the following criteria: 1) rHNSCC/sHNSCC was diagnosed in the presence of a masslike lesion with moderately high (intermediate) signal intensity on T2, low signal intensity on T1, and moderate enhancement after IV gadolinium administration; 2) post-RTH inflammatory edema was diagnosed in the presence of a diffuse lesion with high signal intensity on T2, low signal intensity on T1, and strong postcontrast enhancement; and 3) late RTH fibrosis (post-RTH scar) was diagnosed in the presence of a linear or triangular lesion with very low signal intensity on T2 (similar to or lower than that of muscle), low signal intensity on T1, and faint/absent contrast enhancement.^{1,14,15}

DWI was assessed visually (qualitative DWI) by analyzing the signal intensity on the $b=1000$ image and on the corresponding ADC map. The association of moderately high or very high signal on $b=1000$ and low signal on ADC maps was interpreted as suggesting malignancy. The association of high signal on both $b=1000$ and the ADC map (T2 shinethrough) or low signal on $b=1000$ and high signal on the ADC map was interpreted as post-RTH inflammatory edema, whereas low signal on $b=1000$ and low signal on ADC was regarded as suggesting RTH-induced late fibrosis. Quantitative DWI analysis was based on ADC measurements in benign and malignant lesions, which was performed by drawing small ROIs on multiple slices manually, according to the recommendations in the literature.^{5,16,17} In large, masslike lesions, ROIs were placed on the most hypointense lesion parts on the ADC maps, avoiding necrotic portions.^{5,16}

For combined DWIMRI interpretation, whenever MRI morphology and DWI findings were concordant, the diagnosis of tumor was straightforward. Whenever morphology and DWI were discordant, on the basis of our experience, we applied the following criteria: 1) For lesions with characteristic morphologic features of late fibrosis, morphology outweighed positive DWI

Table 1: Patient and tumor characteristics

	Primary HNSCC	rHNSCC and sHNSCC after Treatment
Total No. of patients		100
Total No. of tumors	103	38 ^a
Female (No.) (%)		22 (22%)
Male (No.) (%)		78 (78%)
Mean age (yr)	59.3 ± 11.3	61.5 ± 11.1
Treatment modalities in pHNSCC (No. of patients)		NA
RTH alone	52	
Operation followed by RTH	48	
Median interval (quartile 1–quartile 3) between end of RTH and rHNSCC/sHNSCC (mo)	NA	14 (4.5–51)
Tumor location (No. of tumor sites) (%)		
Nasopharynx	7 (6.8%)	1 (2.6%)
Oral cavity	22 (21.3%)	13 (34.2%)
Oropharynx	38 (36.9%)	8 (21.1%)
Hypopharynx	12 (11.6%)	5 (13.1%)
Larynx	19 (18.5%)	6 (15.8%)
Paranasal sinuses	4 (3.9%)	2 (5.3%)
Base of the skull	0 (0%)	3 (7.9%)
Unknown primary tumor	1 (1%)	0 (0%)
T classification ^b		
Tx	1 (1%)	0 (0%)
Tis	1 (1%)	1 (2.6%)
T1	14 (13.6%)	3 (7.9%)
T2	30 (29.1%)	7 (18.4%)
T3	25 (24.3%)	6 (15.8%)
T4	32 (31.0%)	21 (55.3%)
N classification ^b		
N0	31 (30.1%)	26 (68.4%)
N1	20 (19.4%)	5 (13.2%)
N2	45 (43.7%)	6 (15.8%)
N3	6 (5.8%)	1 (2.6%)
Nx	1 (1%)	0 (0%)
M classification ^b		
Mx	7 (6.8%)	0 (0%)
M0	94 (91.3%)	36 (94.7%)
M1	2 (1.9%)	2 (5.3%)

Note:—NA indicates not applicable.

^a Thirty-three rHNSCCs and 5 sHNSCCs.

^b Tumor, Node, Metastasis (TNM) classification according to Union for International Cancer Control 2016.²⁸

findings; 2) for lesions with morphology strongly suggesting rHNSCC/sHNSCC, morphology outweighed negative DWI findings; and 3) in the absence of a suspicious lesion at morphology, positive findings on DWI outweighed negative morphology. For the analysis of combined DWIMRI, we recorded the number of instances in which DWI changed the interpretation of morphologic MRI; and on the contrary, we recorded the number of instances, in which MRI changed the interpretation of DWI.

Standard of Reference

The data extracted from the medical records were used for the validation of findings. The distinction between rHNSCC and sHNSCC was made on the basis of the interdisciplinary HN oncology tumor board reports. In rHNSCC, differentiation was made between local recurrence (at the same site as the pHNSCC or in the immediate tumor vicinity) and distant recurrence (eg, base of the skull due to perineural tumor spread). The standard of reference consisted of histology and follow-up. Histology from endoscopic biopsy and/or salvage operation and additional follow-up were available in 50 lesions (37 HNSCCs and 13 benign

lesions), whereas follow-up alone was available in 53 lesions (1 HNSCC and 52 benign lesions). Follow-up included clinical examination and cross-sectional imaging for at least 18 months after MRI. It was used to validate negative results as true-negatives in patients without biopsy and in patients with negative biopsy results. Twenty-nine patients died from disease during the follow-up with a median time interval (25th quartile–75th quartile) between MRI and death of 10 months (4–15 months). The median follow-up in the 71 surviving patients was 25 months (19–33 months).

Statistical Analysis

Comparison between ADCmean values for tumors and benign lesions was performed with a Wilcoxon rank sum test with continuity correction. Receiver operating characteristic analysis of ADCmean values categorized according to the standard of reference was performed. An optimal cutoff value for ADCmean was calculated according to the literature.¹⁸ The diagnostic performance of MRI, DWI, qualitative DWIMRI, and quantitative DWIMRI was assessed in terms of sensitivity, specificity, positive and negative predictive values, and positive and negative likelihood ratios (LR+, LR–). Percentage agreement and Cohen κ coefficients were used to assess interobserver reproducibility, with κ values of 0.21–0.40 indicating fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, almost perfect agreement.¹⁹ Comparison between sensitivities and specificities was performed with the McNemar test, whereas LRs were compared using a regression model approach.²⁰ Statistical analyses were conducted with R3.3.1 (<http://www.R-project.org>).²¹ Statistical tests were 2-sided with a significance level of .05.

RESULTS

Patient and Tumor Characteristics

Patient and tumor characteristics are shown in Table 1. Post-RTH HNSCC (rHNSCC/sHNSCC) was diagnosed in 36 patients in 38 locations (tumor prevalence = 37%; 95% CI, 28%–47%; 2 patients had 2 HNSCCs each). There were 31 local rHNSCCs, 2 distant rHNSCCs (base of the skull via perineural spread), and 5 sHNSCCs in the HN region. Most post-RTH tumors ($n = 33$) occurred in the same anatomic subsite as the pHNSCC ($n = 23$) or at the periphery of the RTH field ($n = 10$). Nevertheless, patients with oropharyngeal pHNSCC developed sHNSCC ($n = 5$) more often than patients with pHNSCC originating in other HN subsites ($P < .001$).

Table 2: Diagnostic performance of morphologic MRI alone, quantitative DWI alone, qualitative DWIMRI, and quantitative DWIMRI with ADCmean < 1.22 × 10⁻³ mm²/s

	Morphologic MRI	Quantitative DWI with ADCmean < 1.22	Morphologic MRI with Qualitative DWI	Morphologic MRI with Quantitative DWI (ADCmean < 1.22)
TP (No.)	30	30	34	35
TN (No.)	52	56	59	62
FP (No.)	13	9	6	3
FN (No.)	8	8	4	3
Sensitivity (%) ^a	78.9	78.9	89.4	92.1
(95% CI)	(65.9–91.9)	(65.9–91.9)	(79.7–99.2)	(83.5–100.0)
Specificity (%) ^b	80.0	86.1	90.8	95.4
(95% CI)	(70.3–89.7)	(74.8–93.1)	(83.7–97.8)	(90.3–100.0)
PPV (%)	69.7	76.9	85.0	92.1
(95% CI)	(56.1–83.5)	(60.3–88.3)	(73.9–96.1)	(83.5–100.0)
NPV (%)	86.6	87.5	93.6	95.4
(95% CI)	(78.1–95.3)	(76.3–94.1)	(87.6–99.6)	(90.2–100.0)
LR+ ^c	3.94	5.70	9.69	19.9
(95% CI)	(2.36–6.59)	(3.04–10.68)	(4.48–20.9)	(6.58–60.5)
LR- ^d	0.26	0.24	0.11	0.08
(95% CI)	(0.14–0.49)	(0.13–0.45)	(0.04–0.29)	(0.03–0.24)

Note:—TP indicates true-positive; TN, true-negative; FP, false-positive; FN, false-negative; PPV, positive predictive value; NPV, negative predictive value.

^a Comparison of sensitivities: MRI vs DWI: *P* = 1; MRI vs qualitative DWIMRI: *P* = .10; MRI vs quantitative DWIMRI: *P* = .025; qualitative DWIMRI vs quantitative DWIMRI: *P* = .31; DWI vs quantitative DWIMRI: *P* = .05.

^b Comparison of specificities: MRI vs DWI: *P* = .34; MRI vs qualitative DWIMRI: *P* = .05; MRI vs quantitative DWIMRI: *P* = .004; qualitative DWIMRI vs quantitative DWIMRI: *P* = .18; DWI vs quantitative DWIMRI: *P* = .03.

^c Comparison of LR+: MRI vs DWI: *P* = .36; MRI vs qualitative DWIMRI: *P* = .03; MRI vs quantitative DWIMRI: *P* = .004; qualitative DWIMRI vs quantitative DWIMRI: *P* = .17; DWI vs quantitative DWIMRI: *P* = .02.

^d Comparison of LR-: MRI vs DWI: *P* = .85; MRI vs qualitative DWIMRI: *P* = .06; MRI vs quantitative DWIMRI: *P* = .01; qualitative DWIMRI vs quantitative DWIMRI: *P* = .24; DWI vs quantitative DWIMRI: *P* = .04.

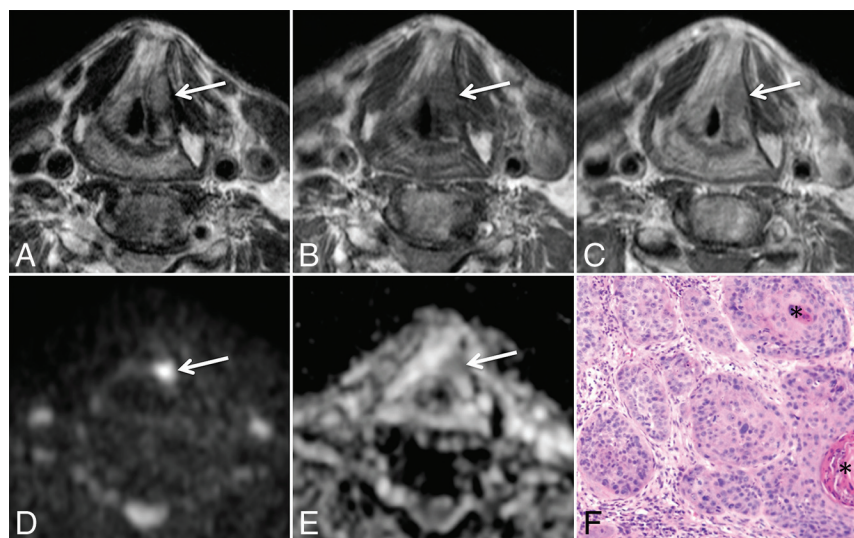


FIG 1. DWIMRI obtained 23 months after RTH for laryngeal squamous cell carcinoma and increasing hoarseness. Axial T2 (A), T1 (B), and contrast-enhanced T1 (C) show an oval lesion (arrows) in the left false cord and left paraglottic space with intermediate signal intensity on T2, low signal intensity on T1, and moderate contrast enhancement highly suspicious for rHNSCC. The *b*=1000 image (D) shows high signal intensity in the lesion. ADC map (E) reveals low signal intensity compatible with restricted diffusion (arrow), further suggesting recurrence (ADCmean = 0.798 × 10⁻³ mm²/s). Endoscopic biopsy confirmed rHNSCC. F, Histology (H&E, original magnification ×100) shows squamous cell carcinoma with areas of densely packed and loosely packed squamous cells of variable size with keratin pearls (asterisk).

Results of Image Interpretation

Morphologic images were of good/acceptable quality in all patients. In 7 patients, DWI showed artifacts in the floor of the mouth due to dental hardware (*n* = 5), tongue base (*n* = 1), or lower neck (*n* = 1). Because both readers considered DWI to be of acceptable quality, these patients were not excluded from analysis.

Evaluations of the 2 readers were concordant in 93/100 patients (96/103 lesions, percentage agreement = 93.2%) with MRI and in 94/100 patients (97/103 lesions, percentage agreement = 94.2%) with qualitative DWIMRI, respectively. The Cohen κ for differentiating tumor from post-RTH lesions with MRI and with qualitative DWIMRI was 0.822 and 0.881, respectively, indicating excellent interobserver reproducibility.¹⁹ Morphologic MRI and DWI were discordant in 28 lesions. The Cohen κ for the 2 readers for combined DWIMRI in this subset of lesions was 0.774 (substantial interobserver reproducibility¹⁹), and the percentage agreement was 89.3%.

Results of the interpretation of MRI and DWIMRI are indicated in Table 2 for the consensus between the 2 readers. Morphologic MRI yielded 30 true-positive evaluations (Fig 1 and On-line Fig 1) and 52 true-negative evaluations (On-line Fig 2). Among the 52 true-negative evaluations, 8 fulfilled the criteria of late fibrosis due to very low signal on T2 (Figs 2 and 3). Both readers correctly identified these lesions. The 8 false-negative evaluations with MRI were caused by the following: abscess from infected osteonecrosis masking recurrence (*n* = 1), in situ tumor (*n* = 1), and T1 tumor (*n* = 1), therefore, too small to be detected by imaging; missed tumor due to poor lesion conspicuity (*n* = 3); and recurrent disease misinterpreted as

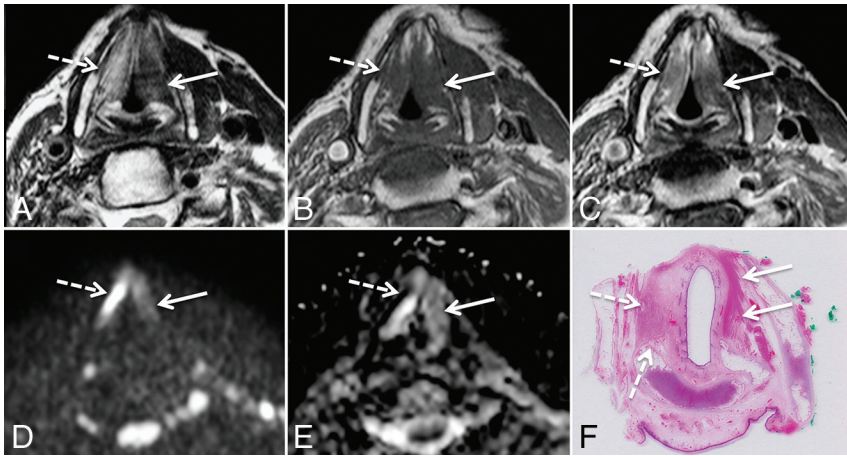


FIG 2. DWIMRI obtained 13 months after RTH and neck dissection for squamous cell carcinoma of the larynx and oropharynx. The patient had massive weight loss, malnutrition, and recurrent aspiration pneumonia. Endoscopy showed intact mucosa and fixed vocal cords bilaterally. Axial T2 (A), T1 (B), and contrast-enhanced T1 (C) show a triangular lesion (arrows) in the left true vocal cord with very low signal intensity on T2, low signal intensity on T1, and faint contrast enhancement suggesting post-RTH late fibrosis. In contrast, the right vocal cord (dashed arrows) displays high signal intensity on T2, low signal on T1, and enhancement. Findings on the right were interpreted as suggesting inflammatory edema. The $b=1000$ image (D) and ADC map (E) reveal no restricted diffusion in the right vocal cord ($ADC_{mean} = 1.643 \times 10^{-3} \text{ mm}^2/\text{s}$) and restricted diffusion with low ADC in the left vocal cord ($ADC_{mean} = 1.006 \times 10^{-3} \text{ mm}^2/\text{s}$). Because the nonfunctional larynx was the cause of malnutrition and recurrent aspiration pneumonia, laryngectomy was performed. F, Corresponding whole-organ histologic slice (H&E) shows extensive muscle fibrosis on the left (arrows) and inflammatory edema with denervation on the right (dashed arrows).

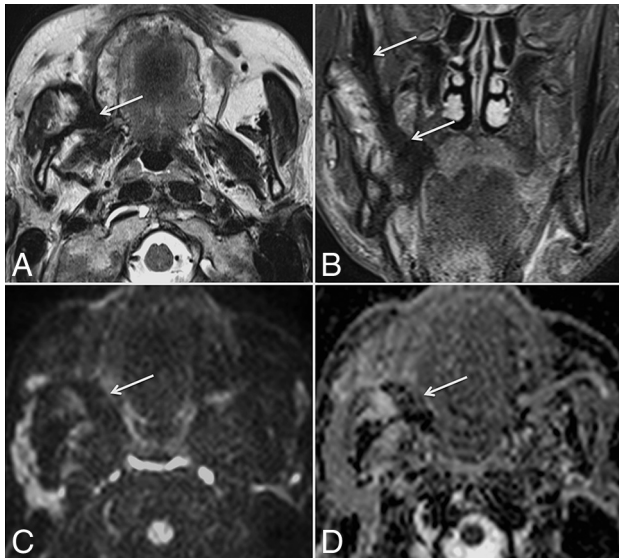


FIG 3. DWIMRI obtained 3 months after RTH and buccopharyngectomy for squamous cell carcinoma of the retromolar trigone. The patient had right reflex otalgia and progressing trismus. Endoscopy could not be performed. Axial T2 (A) and coronal STIR (B) images show a triangular, elongated, strongly hypointense lesion (arrows) on the right. There was no contrast enhancement (not shown). The diagnosis of benign post-RTH late fibrosis was made. The $b=1000$ image (C) reveals low signal. ADC map (D) likewise shows low signal ($ADC_{mean} = 0.731 \times 10^{-3} \text{ mm}^2/\text{s}$). Follow-up at 38 months (not shown) showed no recurrence but progressive scar retraction on MRI.

inflammation ($n = 2$). Thirteen false-positive interpretations with MRI were caused by granulation tissue, which displayed similar signal intensity on T2 and enhancement similar to that of rHNSCC/sHNSCC.

Quantitative DWIMRI analysis revealed that ADC_{mean}

values were significantly lower in rHNSCC/sHNSCC than in benign changes, despite overlap between both groups (mean $ADC_{mean} = 1.097 \pm 0.295 \times 10^{-3} \text{ mm}^2/\text{s}$ versus $1.659 \pm 0.418 \times 10^{-3} \text{ mm}^2/\text{s}$, respectively; $P < .001$). Boxplots for ADC_{mean} of benign-versus-malignant lesions are shown in Fig 4. Boxplots for ADC_{mean} in post-RTH tumors versus late fibrosis and inflammatory edema revealed that late fibrosis had ADC_{mean} values similar to those of malignant tumors (Fig 5). Mean ADC_{mean} was $0.987 \pm 0.264 \times 10^{-3} \text{ mm}^2/\text{s}$ in post-RTH late fibrosis and $1.097 \pm 0.295 \times 10^{-3} \text{ mm}^2/\text{s}$ in tumors, respectively ($P > .05$).

The receiver operating characteristics analysis of ADC measurements (Fig 6) provided an optimal ADC_{mean} threshold of $1.222 \times 10^{-3} \text{ mm}^2/\text{s}$ (sensitivity = 78.9%, specificity = 86.1%, positive predictive value = 76.9%, negative predictive value = 87.5%) for ADC_{mean} values only.

Results for qualitative (visual) DWIMRI and quantitative DWIMRI (with an ADC_{mean} threshold of $< 1.22 \times 10^{-3} \text{ mm}^2/\text{s}$) are shown in Table 2. In comparison with morphologic MRI, qualitative and quantitative DWIMRI enabled the detection of post-RTH HNSCCs with poor morphologic conspicuity in 4 and 5 cases, respectively. Likewise, false-positive MRI interpretations were avoided with qualitative and quantitative DWIMRI in 10 and 11 instances, respectively. On the contrary, in comparison with DWI alone, morphologic MRI enabled avoiding 7 false-positive DWI evaluations due to the characteristic aspect of late fibrosis on T2 and facilitated the detection of 5 recurrent cancers, which were missed on DWI (On-line Fig 3).

Review of the Literature

For the discussion of our results, we have performed a review of the current literature for the detection of post-RTH HNSCC with DWIMRI and have calculated the respective likelihood ratios (On-line Table).

DISCUSSION

The studies demonstrating the utility of DWIMRI for the detection of post-RTH HNSCC (On-line Table) are based on the knowledge that high cellularity of recurrent tumors results in low ADC_{mean} values.^{5-9,11-13} It is speculated that post-RTH tissue changes (inflammatory edema, necrosis, fibrosis) lead to high ADC_{mean} values. However, no study has examined the contribution of each post-RTH tissue change to the overall post-RTH ADC_{mean} . Literature shows that despite sparse references, post-RTH cerebral gliosis may show low ADC_{mean} and an abscess may show restricted diffusion.²² Moreover, rHNSCC with cystic or necrotic components may show high ADC_{mean} . Despite this di-

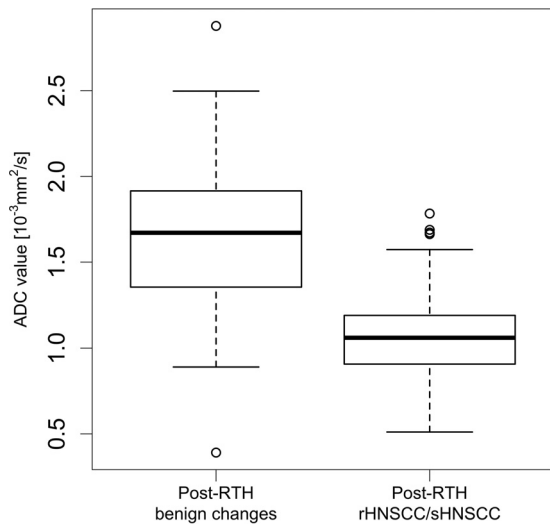


FIG 4. Box-and-whisker plots of ADCmean values in patients with post-RTH changes and post-RTH tumors. The *horizontal lines* represent the median values, and the bottom and the top of the box represent the 25th and 75th quartiles, respectively. Median ADCmean (25th–75th quartiles) for rHNSCC/sHNSCC = $1.061 (0.907\text{--}1.191) \times 10^{-3} \text{ mm}^2/\text{s}$. Median ADCmean (25th–75th quartiles) for post-RTH changes (late fibrosis and inflammatory edema together) = $1.671 (1.3355\text{--}1.915) \times 10^{-3} \text{ mm}^2/\text{s}$.

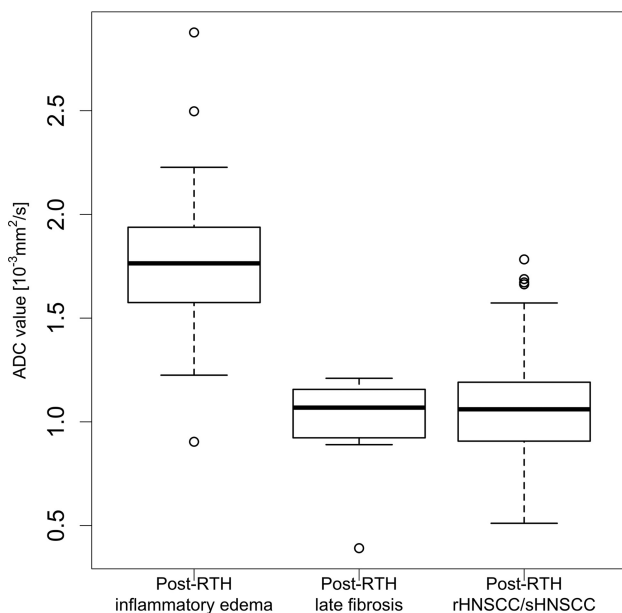


FIG 5. Box-and-whisker plots of ADCmean values in patients with post-RTH inflammatory edema, late fibrosis, and post-RTH HNSCCs. The *horizontal lines* represent the median values; the bottom and the top of the box represent the 25th and 75th quartiles, respectively. Median ADCmean (25th–75th quartiles) for rHNSCC/sHNSCC = $1.061 (0.907\text{--}1.191) \times 10^{-3} \text{ mm}^2/\text{s}$. Median ADCmean (25th–75th quartiles) for post-RTH inflammatory edema = $1.764 (1.575\text{--}1.938) \times 10^{-3} \text{ mm}^2/\text{s}$. Median ADCmean (25th–75th quartiles) for late fibrosis/mature scar post-RTH = $1.068 (0.939\text{--}1.152) \times 10^{-3} \text{ mm}^2/\text{s}$. There was no statistically significant difference between ADCmean in late fibrosis and rHNSCC/sHNSCC ($P > .05$). However, there was a significant difference between ADCmean in inflammatory post-RTH edema and late fibrosis ($P < .05$).

versity, in all aforementioned studies, the ADCs of various post-RTH changes were grouped together as “post-RTH ADCmean” and were compared with the ADCs of rHNSCC. This comparison

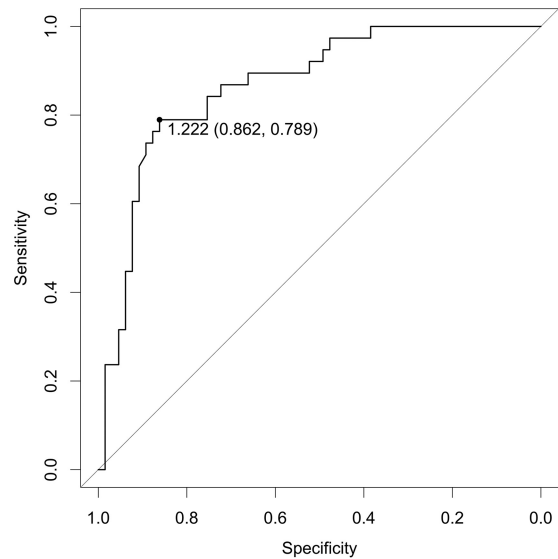


FIG 6. Receiver operating characteristic curve for the quantitative analysis of ADCmean values showing the area under the curve of 0.8678. A threshold of $\text{ADC} = 1.222 \times 10^{-3} \text{ mm}^2/\text{s}$ was found (see description in the text). This threshold yielded a sensitivity of 78.9%, a specificity of 86.2%, and an accuracy of 83.5%.

showed that the ADCmean values of the 2 groups substantially overlapped. Despite the use of similar DWI sequences, the ADC thresholds obtained by these studies for differentiating rHNSCC from post-RTH changes varied from 0.96 to $1.46 \times 10^{-3} \text{ mm}^2/\text{s}$.^{5-8,11-13}

Tshering Vogel et al⁵ have shown that major overlap of ADC values limits the ability of quantitative DWI to differentiate rHNSCC from post-RTH changes. In their study, combined morphologic MRI and qualitative DWI could detect rHNSCC more accurately than combined morphologic MRI and quantitative DWI. However, the criteria used for morphologic analysis were neither defined nor analyzed. In our study, there was no significant difference between the ADCmean of late fibrosis and rHNSCC/sHNSCC ($P > .05$). However, there was a significant difference between the ADCmean of post-RTH edema and late fibrosis ($P < .05$), with virtually no overlap between these 2 values. This finding highlights the drawback of grouping ADC values of all post-RTH changes together.

Both Abdel Razeq et al⁷ and Vaid et al¹¹ reported false-positive results with DWIMRI due to late fibrosis. Other studies have not discussed the causes of false-positive cases. Although the T2 appearance of late fibrosis has been briefly addressed by some authors, its diagnostic utility as a complementary tool to DWI and its impact on the diagnostic performance have not been analyzed.^{1,14,23} Because late fibrosis is mainly composed of densely packed collagen, ADCs tend to be low (T2 blackout effect) as opposed to RTH-induced inflammatory edema, which tends to have high ADCs. In the current study, morphologic MRI, mainly T2, enabled accurate diagnosis of late fibrosis, which otherwise may mimic rHNSCC on DWI.

Previous investigators have not compared the diagnostic performance of DWI with that of morphologic MRI. Data of the current series show that morphologic MRI with clearly defined criteria has a similar diagnostic performance to DWI alone; how-

ever, the combination of the 2 yields superior results. Careful analysis of signal intensities on morphologic MRI contributes significantly to an improved specificity of DWIMRI, whereas the overall effect on sensitivity is less pronounced.

The wide variability in sensitivity, specificity, and positive and negative predictive values of DWIMRI demonstrated by the aforementioned studies can be explained by several factors: histologic heterogeneity (HNSCC only versus HNSCC mixed with other tumor types), local recurrence only versus combined local and nodal recurrence, small sample size, varying tumor prevalence, and differences in ROI sampling.^{5-8,11-13} Although sensitivity and specificity are not influenced by disease prevalence, they can vary depending on the disease spectrum in the study population. Disease prevalence has an important impact on the positive predictive value and a slightly weaker influence on the negative predictive value. Therefore, the use of a likelihood ratio as a measure of test accuracy overcomes the drawback of other metrics because it does not depend on disease prevalence.²⁴ Consequently, the LRs from 1 study are applicable to other studies. A LR+ of >10 indicates a good diagnostic test for ruling in the diagnosis, whereas an LR- of <0.1 indicates a good test for ruling out the diagnosis.²⁴ Results of our study show that quantitative DWIMRI was clearly superior to morphologic MRI and DWI alone because the LR+ and LR- (19.9 and 0.08, respectively) fulfilled the criteria of a good diagnostic test. Comparison with the literature further suggests that except for Vandecaveye et al,⁶ the LR+ and LR- of quantitative DWIMRI were suboptimal in all published series; the values were similar to those of DWI alone and morphologic MRI alone in the current study.

To the best of our knowledge, except for Vaid et al,¹¹ none of the previous authors tested the interobserver reproducibility of their results regarding the assessment of post-RTH cases. In our study, the κ values for MRI and DWIMRI were >81%, suggesting almost perfect interobserver agreement.¹⁹ Therefore, our results are reproducible, provided that experienced readers perform the image analysis. Although the guidelines for the interpretation of interrater reliability proposed by Landis and Koch¹⁹ are widely used, several authors have questioned judgments about acceptable κ values in medical research.²⁵ The κ statistic was initially designed to account for raters guessing on scores. Despite this advantage, the κ statistic may substantially lower the estimate of agreement. Therefore, reporting both κ and percentage agreement may represent a good compromise.²⁵

Our study has some limitations. It is a retrospective study of patients from a single hospital, which could have created a selection bias. Exclusion of patients who were followed up with CT and PET/CT and those with nodal recurrence only was consistent with the aim of our study and was unavoidable. To maintain homogeneity of results, we included only patients scanned on a 1.5T MRI scanner. These factors could have further added to selection bias. Another study limitation is the lack of quantitative T2 and contrast-enhanced T1 signal analysis. Although categorizing signal intensity as high, intermediate, and low may appear somewhat subjective, this visual approach is widely used in HN routine and for the assessment of tumors in other organs. Quantification of T2 and T1 requires the additional acquisition of mapping sequences to calculate relaxation maps. Quantitative information can addi-

tionally be extracted with texture/histogram analysis. This post-processing tool can be applied retrospectively to all images acquired in clinical routine.^{26,27} However, lack of standardized software, evolving research, and variable numeric algorithms currently limit the use of these new promising tools in clinical routine.

CONCLUSIONS

The combination of precise morphologic MRI criteria and quantitative DWI with an ADCmean < 1.22×10^{-3} mm² has an excellent interobserver reproducibility and yields likelihood ratios of a good diagnostic test to rule in and out post-RTH rHNSCC/shHNSCC. Our study shows that morphologic MRI criteria and DWI are complementary and contribute differently to the overall diagnostic performance of combined DWIMRI.

REFERENCES

1. Varoquaux A, Rager O, Dulguerov P, et al. **Diffusion-weighted and PET/MR imaging after radiation therapy for malignant head and neck tumors.** *Radiographics* 2015;35:1502-27 CrossRef Medline
2. de Bree R, van der Putten L, Brouwer J, et al. **Detection of locoregional recurrent head and neck cancer after (chemo)radiotherapy using modern imaging.** *Oral Oncol* 2009;45:386-93 CrossRef Medline
3. King AD, Thoeny HC. **Functional MRI for the prediction of treatment response in head and neck squamous cell carcinoma: potential and limitations.** *Cancer Imaging* 2016;16:23 CrossRef Medline
4. Brouwer J, Bodar EJ, De Bree R, et al. **Detecting recurrent laryngeal carcinoma after radiotherapy: room for improvement.** *Eur Arch Otorhinolaryngol* 2004;261:417-22 Medline
5. Tshering Vogel DW, Zbaeren P, Geretschlaeger A, et al. **Diffusion-weighted MR imaging including bi-exponential fitting for the detection of recurrent or residual tumour after (chemo)radiotherapy for laryngeal and hypopharyngeal cancers.** *Eur Radiol* 2013;23:562-69 CrossRef Medline
6. Vandecaveye V, De Keyser F, Nuyts S, et al. **Detection of head and neck squamous cell carcinoma with diffusion weighted MRI after (chemo)radiotherapy: correlation between radiologic and histopathologic findings.** *Int J Radiat Oncol Biol Phys* 2007;67:960-71 CrossRef Medline
7. Abdel Razek AA, Kandeel AY, Soliman N, et al. **Role of diffusion-weighted echo-planar MR imaging in differentiation of residual or recurrent head and neck tumors and posttreatment changes.** *AJNR Am J Neuroradiol* 2007;28:1146-52 CrossRef Medline
8. Hwang I, Choi SH, Kim YJ, et al. **Differentiation of recurrent tumor and posttreatment changes in head and neck squamous cell carcinoma: application of high b-value diffusion-weighted imaging.** *AJNR Am J Neuroradiol* 2013;34:2343-48 CrossRef Medline
9. King AD, Mo FK, Yu KH, et al. **Squamous cell carcinoma of the head and neck: diffusion-weighted MR imaging for prediction and monitoring of treatment response.** *Eur Radiol* 2010;20:2213-20 CrossRef Medline
10. Mukherji SK, Wolf GT. **Evaluation of head and neck squamous cell carcinoma after treatment.** *AJNR Am J Neuroradiol* 2003;24:1743-46 Medline
11. Vaid S, Chandorkar A, Atre A, et al. **Differentiating recurrent tumours from post-treatment changes in head and neck cancers: does diffusion-weighted MRI solve the eternal dilemma?** *Clin Radiol* 2017;72:74-83 CrossRef Medline
12. Desouky SS, AboSeif SS, Shama SS, et al. **Role of dynamic contrast enhanced and diffusion weighted MRI in the differentiation between post treatment changes and recurrent laryngeal cancers.** *Egypt J Radiol Nucl Med* 2015;46:379-89 CrossRef
13. Gouhar GK, El-Hariri MA. **Feasibility of diffusion weighted MR im-**

- aging in differentiating recurrent laryngeal carcinoma from radionecrosis. *Egypt J Radiol Nucl Med* 2011;42:169–75 CrossRef
14. King AD, Keung CK, Yu KH, et al. **T2-weighted MR imaging early after chemoradiotherapy to evaluate treatment response in head and neck squamous cell carcinoma.** *AJNR Am J Neuroradiol* 2013;34:1237–41 CrossRef Medline
 15. Becker M, Varoquaux AD, Combescure C, et al. **Local recurrence of squamous cell carcinoma of the head and neck after radio(chemo)therapy: diagnostic performance of FDG-PET/MRI with diffusion-weighted sequences.** *Eur Radiol* 2018;28:651–63 CrossRef Medline
 16. Thoeny HC, De Keyzer F, King AD. **Diffusion-weighted MR imaging in the head and neck.** *Radiology* 2012;263:19–32 CrossRef Medline
 17. 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 CrossRef Medline
 18. Perkins NJ, Schisterman EF. **The inconsistency of “optimal” cut-points obtained using two criteria based on the receiver operating characteristic curve.** *Am J Epidemiol* 2006;163:670–75 CrossRef Medline
 19. Landis JR, Koch GG. **The measurement of observer agreement for categorical data.** *Biometrics* 1977;33:159–74 CrossRef Medline
 20. Gu W, Pepe MS. **Estimating the diagnostic likelihood ratio of a continuous marker.** *Biostatistics* 2011;12:87–101 CrossRef Medline
 21. R Core Team. *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing; 2017
 22. Hein PA, Eskey CJ, Dunn JF, et al. **Diffusion-weighted imaging in the follow-up of treated high-grade gliomas: tumor recurrence versus radiation injury.** *AJNR Am J Neuroradiol* 2004;25:201–09 Medline
 23. Farina D, Borghesi A, Botturi E, et al. **Treatment monitoring of paranasal sinus tumors by magnetic resonance imaging.** *Cancer Imaging* 2010;10:183–93 CrossRef Medline
 24. Šimundić AM. **Measures of diagnostic accuracy: basic definitions.** *EJIFCC* 2009;19:203–11 Medline
 25. McHugh ML. **Interrater reliability: the kappa statistic.** *Biochem Med* 2012;22:276–82 Medline
 26. Kuno H, Qureshi MM, Chapman MN, et al. **CT texture analysis potentially predicts local failure in head and neck squamous cell carcinoma treated with chemoradiotherapy.** *AJNR Am J Neuroradiol* 2017;38:2334–40 CrossRef Medline
 27. de Perrot T, Lenoir V, Domingo Ayllon M, et al. **Apparent diffusion coefficient histograms of human papillomavirus-positive and human papillomavirus-negative head and neck squamous cell carcinoma: assessment of tumor heterogeneity and comparison with histopathology.** *AJNR Am J Neuroradiol* 2017;38:2153–60 CrossRef Medline
 28. Brierley JD, Gospodarowicz MK, Wittekind C. *TNM Classification of Malignant Tumours.* 8th ed. Oxford: Wiley-Blackwell; 2017