POLISH
JOURNAL *of* **RADIOLOGY**

http://www.polradiol.com

Original paper

External validation of the Brain Tumour Reporting and Data System (BT-RADS) in the multidisciplinary management of post-treatment gliomas

Kamaxi Hitendrakumar Trivedi[1,A,B,C,D,E,F], Amrita Guha[1,A,B,C,D,E], Meenakshi Thakur[1,D,E], Abhishek Mahajan[1,2,A,D,E], Pallavi Bhole[1,C], Tejpal Gupta[1,E]

[1]Tata Memorial Centre, Homi Bhabha National Institute, Mumbai, India
[2]The Clatterbridge Cancer Centre, University of Liverpool, Liverpool, United Kingdom

Abstract

Purpose: To independently and externally validate the Brain Tumour Reporting and Data System (BT-RADS) for post-treatment gliomas and assess interobserver variability.

Material and methods: In this retrospective observational study, consecutive MRIs of 100 post-treatment glioma patients were reviewed by two independent radiologists (RD1 and RD2) and assigned a BT-RADS score. Inter-observer agreement statistics were determined by kappa statistics. The BT-RADS-linked management recommendations per score were compared with the multidisciplinary meeting (MDM) decisions.

Results: The overall agreement rate between RD1 and RD2 was 62.7% (κ = 0.67). The agreement rate between RD1 and consensus was 83.3% (κ = 0.85), while the agreement between RD2 and consensus was 69.3% (κ = 0.79). Among the radiologists, agreement was highest for score 2 and lowest for score 3b. There was a 97.9% agreement between BT-RADS-linked management recommendations and MDM decisions.

Conclusions: BT-RADS scoring led to improved consistency, and standardised language in the structured MRI reporting of post-treatment brain tumours. It demonstrated good overall agreement among the reporting radiologists at both extremes; however, variation rates increased in the middle part of the spectrum. The interpretation categories linked to management decisions showed a near-perfect match with MDM decisions.

Key words: post-treatment glioma, BT-RADS, structured reporting, neuroradiology, external validation.

Introduction

Gliomas constitute a group of heterogenous malignant primary brain tumours, with a median survival duration that spans from 7 years for low-grade gliomas to 18 months for high-grade gliomas, such as glioblastoma [1,2]. Standard therapy involves gross total/near total resection, conformal radiotherapy and alkylating agents like temozolomide (TMZ) [3].

Magnetic resonance (MR) imaging has been instrumental in the surveillance of post-treatment gliomas; however, the complex clinical course, varied patterns of presentation, overlapping features of tumour progression and treatment effect, and the heterogeneous nature of each case present challenges in the interpretation and reporting of the findings [4]. Several criteria for response assessment in gliomas have been proposed, including Macdonald and Response Assessment in Neuro-Oncology criteria [5-7]. These systems have limited utility in clinical

practice due to their complexity, lack of objectivity, and time-intensive nature. In order to address this clinical gap, a structured reporting system, the Brain Tumour Reporting and Data System (BT-RADS), was proposed by the Emory neuroradiology group, wherein MRI studies are interpreted and reported systematically to provide objective clarity to the referring clinicians in line with the current standard management recommendations [8]. Our objective was to independently and externally validate the BT-RADS scoring system, and evaluate its interobserver variability and accuracy in directing management. We also attempted to investigate whether the BT-RADS score has any predictive value for survival outcomes.

Material and methods

Patient selection and interpretation

This retrospective study was approved by our institutional review board (project no. 900621), which waived the requirement for written informed consent. We included post-treatment glioma patients who underwent contrast-enhanced MRI between November 2018 and June 2019. We curated the following information from the patient's chart in electronic medical records (EMR) and the Picture Archiving and Communication System (PACS): patient demographics, diagnosis and follow-up details, treatment history, and imaging parameters for two consecutive MRIs. Two-year follow-up data were collected.

MRI protocol

All patients underwent MR brain examination using dedicated phased-array head coils on one of the following scanners: Ingenia, Philips (1.5T), Signa, or GE (1.5T and 3T). A standardised MR brain tumour protocol was

followed, and the pulse sequences that were used are summarised in Table 1.

Image analyses

Image interpretation was performed independently by two neuro-radiologists (RD1 with 10 years and RD2 with 4 years of neuro-oncologic imaging experience) on a dedicated clinical workstation using picture archiving and communication systems (PACS). The readers were aware that all patients were post-treatment glioma cases but were blinded to all other outcomes. The readers assigned a BT-RADS score to each patient after assessing two consecutive post-treatment MRIs.

Eligibility criteria and scoring system

We included all postoperative glioma patients undergoing adjuvant therapy (radiation therapy ± TMZ) and excluded paediatric (≤ 18 years) brain tumours and patients without histological diagnosis as per electronic medical records. The study design is depicted in the graphic abstract shown in Figure 1. The adapted scoring lexicon system and its definitions are provided under supplementary Tables S1 and S2. Any discordance between the two assigned scores was then resolved through mutual discussion, and a final "consensus" score was assigned. The consensus score was considered as the ground truth for all analysis. Each of these scores was then correlated with the multidisciplinary meeting (MDM), and agreement statistics were derived. We defined "clinically significant" observer variability in cases where there was a substantial difference in the management guidelines linked to the scores assigned by the 2 readers separately for the same patient. We grouped categories that indicate tumour progression (3b, 3c, and 4) and the remaining categories (1-3a) for assessment of survival outcomes.

Table 1. MRI imaging parameters and sequences utilised in our study

Sequence		Planes	Slice thickness (mm)	Spacing (mm)	TR (ms)	TE (ms)	Matrix
Precontrast T1-weighted spin-echo images		Axial	3	1	600-700	20	320 × 256
		Sagittal	5	0.7-1.5			320 × 192
T2-weighted fast spin-echo		Axial	3	1	2700	100	320 × 256
		Coronal	3-5	1-1.5			320 × 256
T2 FLAIR*	Inversion time: 2200 ms	Axial	3	0.6-1	9000	120	256 × 224
Gradient-echo (GRE) images		Axial	3-5	1	570	30	288 × 224
Diffusion- weighted images	b-values: 0, 50, 1000 s/mm²	Axial	3-5	0-1	8300	70	96 × 96
Postcontrast T1-weighted spin-echo images		Axial	3	1	600-700	20	320 × 192
		Sagittal	5	0.5-1			320 × 192
		Coronal	5	1-1.5			320 × 192
3D FSPGR		Axial	1-2	0	6	4	256 × 224

TR – repetition time, TE – echo time, FLAIR – fluid attenuation and inversion recovery, FSPGR – fast spoiled gradient echo
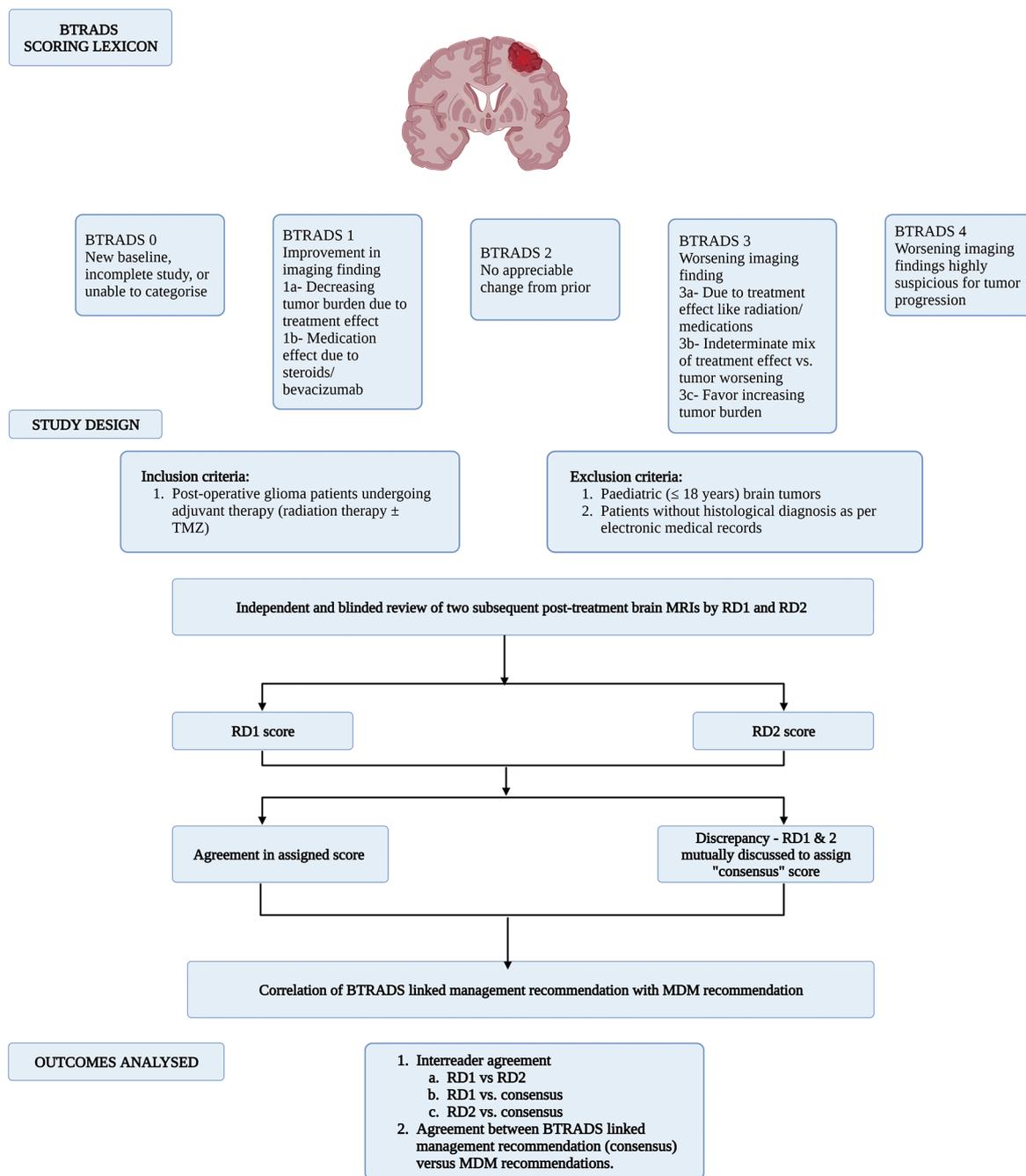
Figure 1. Graphic abstract depicting the study design [8]

Statistical analysis

Data were analysed using IBM SPSS v25. Interobserver agreement was categorised as small for a kappa (κ) value of 0.01-0.20, fair for a κ value of 0.21-0.40, moderate for a κ value of 0.41-0.60, substantial for a κ value of 0.61-0.80, and almost perfect for a κ value of 0.81-1.00 [9]. The overall survival among the BT-RADS scoring categories was assessed using Kaplan-Meier analysis. Death was considered an event, and the time until death/last follow-up was calculated from the date of the subsequent MRI used for assigning scores within the study duration.

Results

Demographics

Table 2 depicts the basic demographic details of our study cohort. The observed frequencies of the scores, listed in descending order, were as follows: BT-RADS 2 in 67 cases, BT-RADS 1 in 9 cases, BT-RADS 3c in 8 cases, BT-RADS 3a in 6 cases, BT-RADS 4 in 5 cases, BT-RADS 0 in 4 cases, and one case of BT-RADS 3b.

Agreement between radiologists and the multidisciplinary meeting decision

Kappa statistics showed low interobserver variability in BT-RADS scores, both between RD1 and RD2 and between each radiologist and the consensus. The overall agreement rate between RD1 and RD2 was 62.7%, with a κ of 0.67. The agreement rate between RD1 and the consensus was 83.3%, with a κ value of 0.85, while the agreement between RD2 and the consensus was 69.3%, with a κ value of 0.79 (Table 3).

Table 2. Demographics and BT-RADS score distribution

Factor	
Age (years), median (range)	39.9 (21-70)
Gender, male/female (ratio)	65/35 (1.9/1)
Glioma type, *n* (%)	
Diffuse astrocytoma	2 (2)
Anaplastic astrocytoma	38 (38)
Diffuse ODG	6 (6)
Anaplastic ODG	19 (19)
Glioblastoma	35 (35)
WHO grade, *n* (%)	
I	0
II	13 (13)
III	51 (51)
IV	36 (36)
BT-RADS score, *n* (%)	
0	4 (4)
1	9 (9)
2	67 (67)
3A	6 (6)
3B	1 (1)
3C	8 (8)
4	5 (5)

Table 3. Kappa statistics agreement amongst RD1, RD2, and consensus

	RD1 vs. RD2*	RD1 vs. Consensus*	RD2 vs. Consensus*
Score 1a	70	70	75
Score 2	92.4	97	97
Score 3a	66.7	83.3	71.4
Score 3b	0	33.3	0
Score 3c	50	100	66.7
Score 4	60	100	75
Overall (average)	62.7	83.3	69.3
κ value (*p*-value)	0.67 ± 0.06 (< 0.001)	0.85 ± 0.04 (< 0.001)	0.79 ± 0.05 (< 0.001)

*All the numbers in the table represent percentages

Table 4. Variation in assigned scores by RD1, RD2, and consensus

		RD2						Total
		1	2	3a	3b	3c	4	
RD1	1	7	2	1	0	0	0	10
	2	5	61	0	0	0	0	66
	3a	0	2	4	0	0	0	6
	3b	0	0	1	0	1	1	3
	3c	0	1	1	1	3	0	6
	4	0	0	0	0	2	3	5
Total		12	66	7	1	6	4	96

		Consensus						Total
		1	2	3a	3b	3c	4	
RD1	1	7	2	1	0	0	0	10
	2	2	61	0	0	0	0	66
	3a	0	1	4	0	0	0	6
	3b	0	0	0	0	2	1	3
	3c	0	1	1	1	6	0	6
	4	0	0	0	0	0	3	5
Total		9	67	6	1	8	5	96

		Consensus						Total
		1	2	3a	3b	3c	4	
RD2	1	9	3	0	0	0	0	12
	2	0	64	1	0	1	0	66
	3a	0	0	5	1	1	0	7
	3b	0	0	0	0	1	0	1
	3c	0	0	0	0	4	2	6
	4	0	0	0	0	1	3	4
Total		9	67	6	1	8	5	96

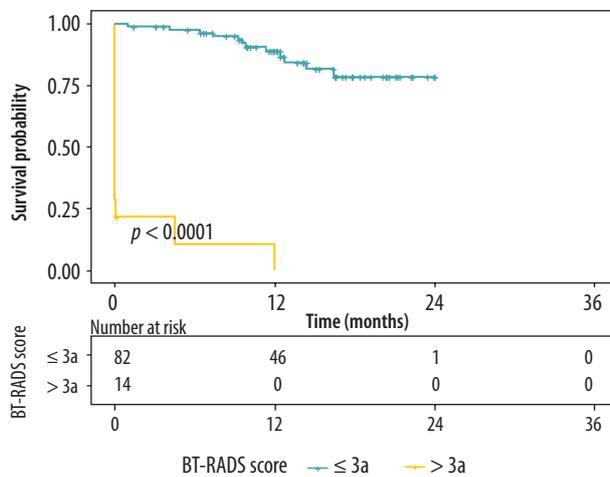**Figure 2.** Kaplan-Meier overall survival analysis curve of BT-RADS ≤ 3a and BT-RADS > 3a in post-treatment glioma patients

Among the radiologists, the highest agreement was observed for score 2, and no concordance was observed for score 3b ($n = 1$) between consensus and both RD1 and RD2. When comparing individual reader findings with the consensus score, there was no clinically significant difference in the assigned scores between RD1 and the consensus. However, RD1 and RD2 assigned differing scores to 3 patients, which was clinically significant (Table 4). As illustrated in Figures 3-5, these 3 patients probably represent true interobserver variability, which can be attributed to the complex nature of post-treatment glioma imaging.

The agreement between Joint Clinic (JC) management decision and BT-RADS-linked management recommendation for each score (consensus BT-RADS score) was 97.9% (in 94 out of 96 cases, because 4 cases were scored 0), with only 2.1% showing disagreement.

Correlation of BT-RADS scoring with overall survival (OS)

Patients were followed up for a median duration of 13.5 (12.2-16.4) months. Categories indicative of tumour progression (3b, 3c, and 4) were grouped together for comparison with the remaining categories (1-3a).

Overall, the one-year survival rate was 87.1% (95% CI: 80.5-94.2) including all patients under surveillance following recent MRI ($p < 0.001$). The probability of 12-month survival with a score $\leq$ 3a was 94.8% (95% CI: 89.9-99.8), whereas for scores 3b or higher it was only 31.5% (95% CI: 4.0-59.0). This disparity in survival rates was statistically significant ($p < 0.001$) (Table 5 and Figure 2).
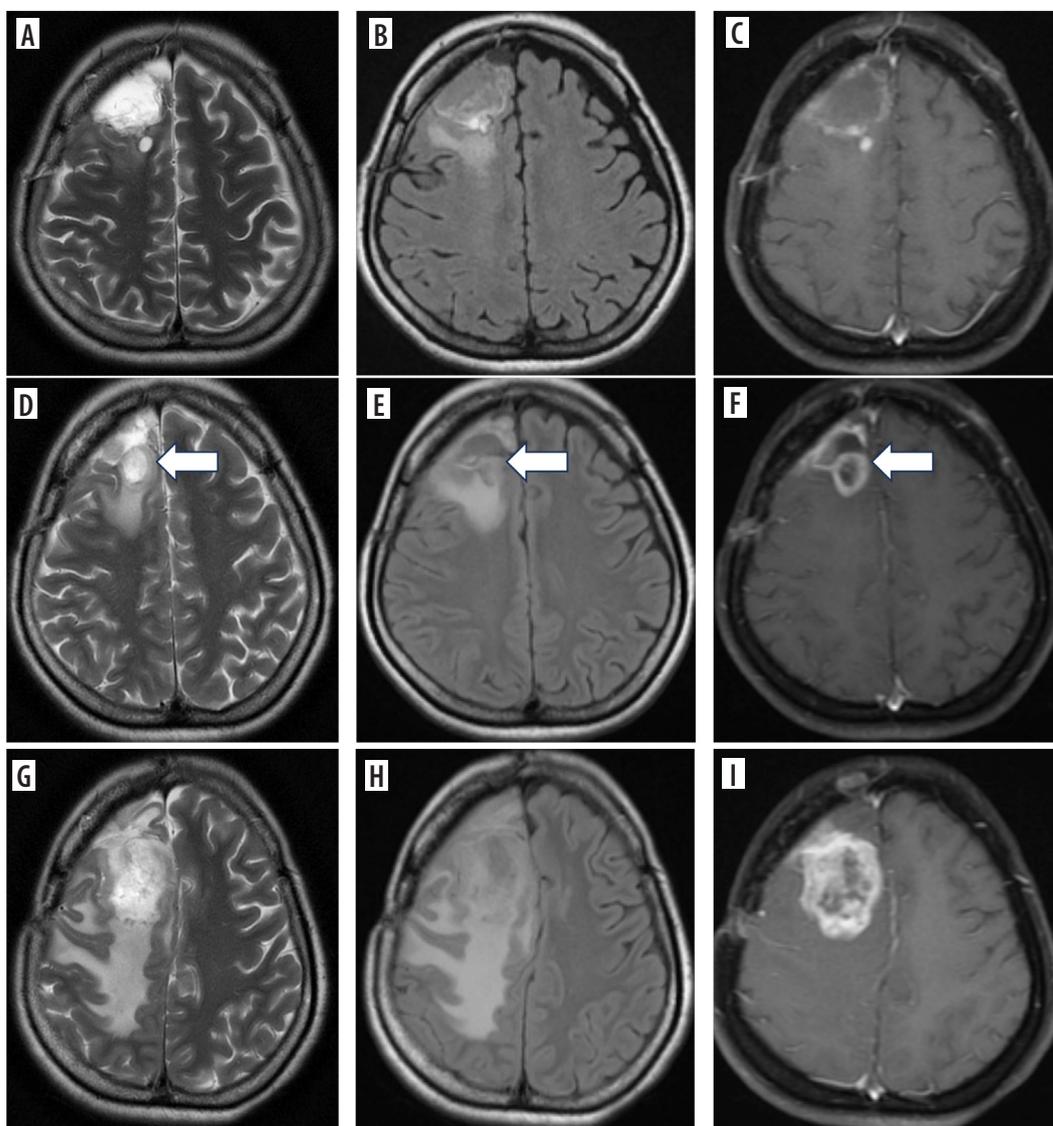

Figure 3. Post-treatment changes vs. recurrence, BT-RADS 3a vs. 3c. A 32-year-old male with glioblastoma (WHO grade IV). Baseline post-surgery (**A-C**) MRI reveals no apparent residual tumour. Follow-up MRI at 2 months after adjuvant radiotherapy (**D-F**) shows a new onset rim enhancing T2/FLAIR hyperintense lesion (white arrows) in the right frontal lobe. Despite being performed within 90 days of radiotherapy, the BT-RADS score was assigned as 3c, indicative of potential recurrence, rather than the expected 3a as per guidelines. Follow-up MRI 7 months after radiotherapy (**G-I**) shows significant increase in the enhancing lesion with mass effect, validating the previously assigned score 3c
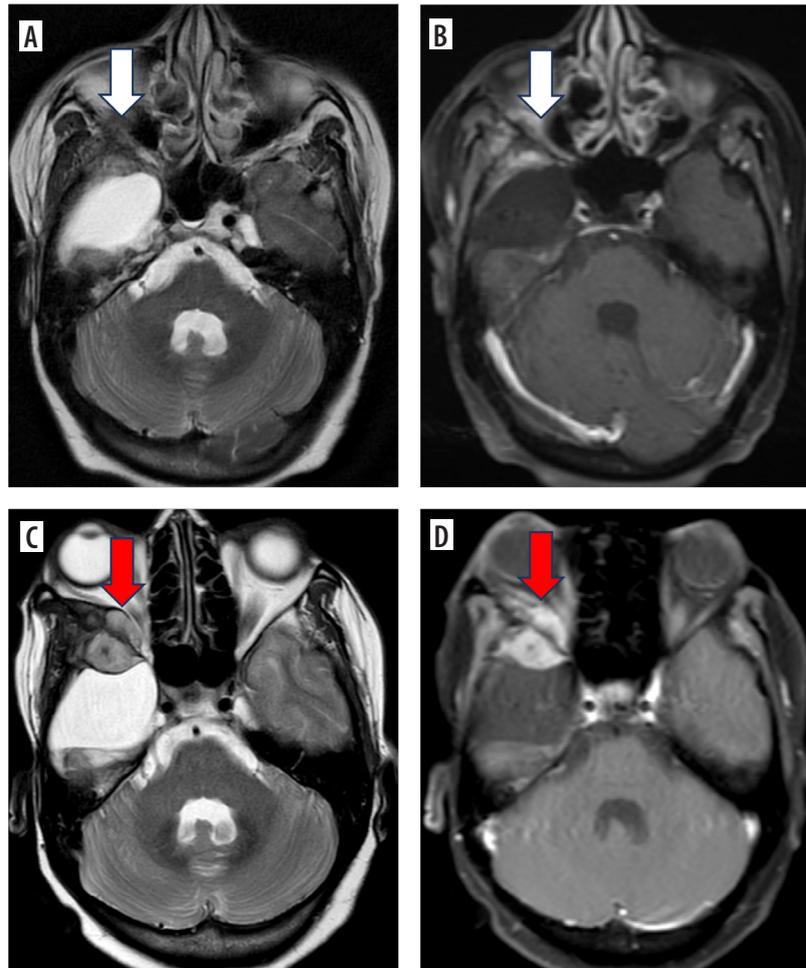
Figure 4. Recurrence vs. radiation induced bone tumour. A 40-year-old female with oligodendroglioma (WHO grade III). Follow-up post-treatment imaging (**A, B**) shows a homogenously enhancing lesion (white arrows) along the anterior temporal lobe, involving the sphenoid wing. Subsequent imaging (**C, D**) after 6 months reveals a substantial increase in the lesion, now popping into the right orbit (red arrows); RD1 considered this as disease progression and scored BT-RADS 3c, while RD2 was of the opinion that this represented an incidental finding of radiation induced bone tumour along the sphenoid bone with no real glioma progression along the tumour bed, and scored BT-RADS 2

Table 5. One-year probability of overall survival in univariate Cox regression analysis

		Overall survival			
		Number	Event	One year probability	*p*-value
Overall		100	15	87.1 (80.5, 94.0)	< 0.001
Final score	≤ 3A	82	6	94.8 (89.9, 99.8)	
	> 3A	14	9	31.5 (4.0, 59.0)	

Discussion

This study is the first in literature to externally and independently validate the use of the BT-RADS scoring system in response assessment of post-treatment gliomas. In our evaluation, a substantial agreement was observed between the two reporting radiologists, yielding a κ value of 62.9% (κ = 0.67) for RD1 and RD2. Furthermore, there was a robust 97.9% consensus between the recommendations from MDM and BT-RADS score-linked management recommendation. Similar results were reported by Cooper *et al.*: an overall agreement of 82.2% was reported between radiologist 1 and radiologist 2, and 79.0% between the initial review and the consensus of the tumour board [10].

Interobserver variability was lowest for scores ≤ 3a, while it was highest for score 3b, followed by 3c. It again dropped for score 4, indicating low variability. This observation highlights that the BT-RADS scoring system functioned well at the extreme scores but with relative ambiguity at the mid-range score, i.e., score 3; assigning this score was intuitive and based on experience, considering the heterogenous nature of the post-treatment imaging and difficulties in differentiating the subsets of pseudoprogression from true progression. This suggests that the scoring system may not be completely objective and may require further refinement or training to improve consistency. The concordance between JC management decisions and BT-RADS-linked management recommendations for each score was 97.9%, indicating a high level of accuracy in this scoring system and underscoring its potential for clinical translation.
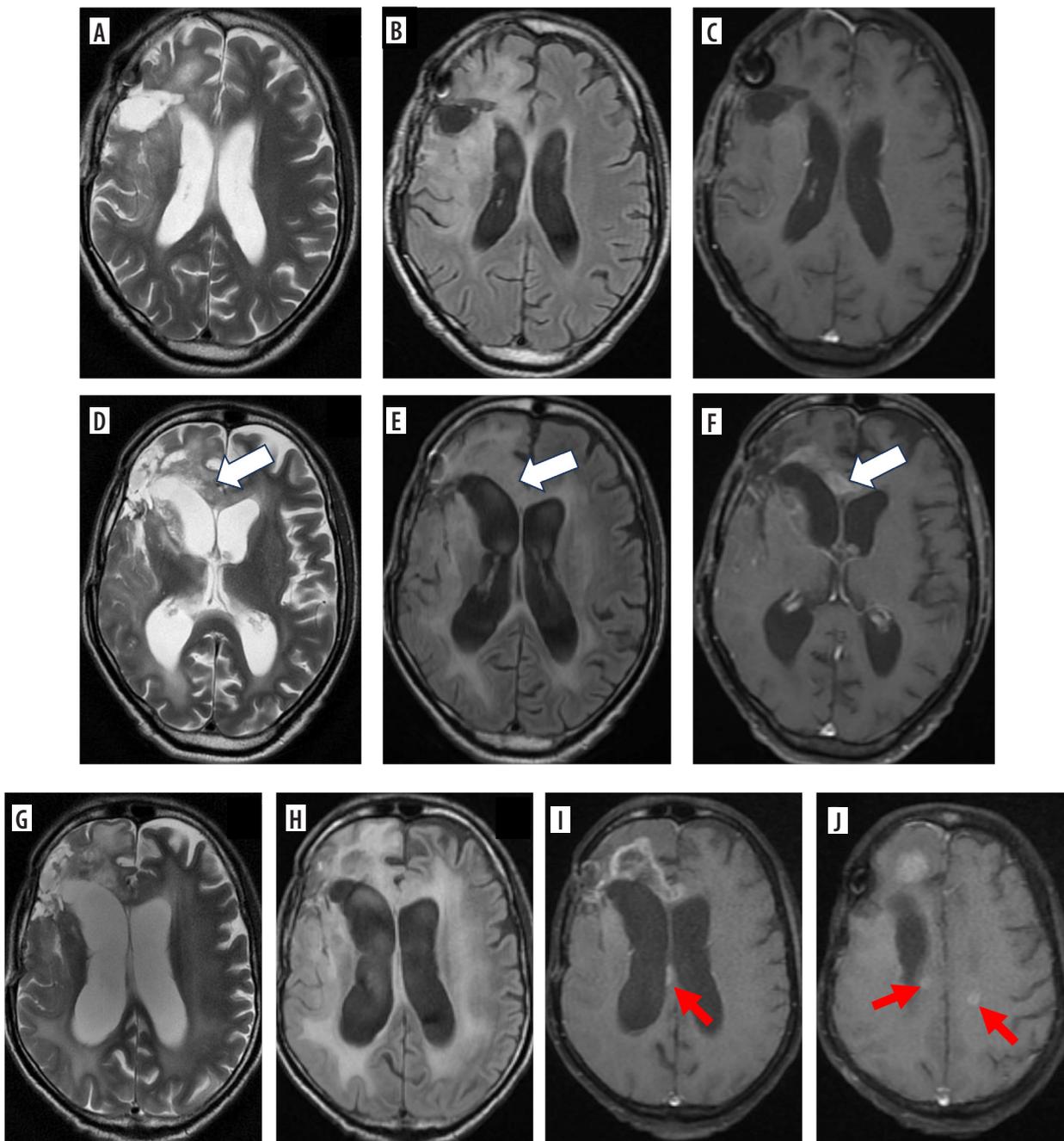
Figure 5. A) Treatment related changes vs. recurrence, BT-RADS 3a vs. 3c. A 45-year-old male with anaplastic astrocytoma (WHO grade III). Comparison of baseline post-resection imaging (**A-C**) and post-treatment imaging within 90 days of radiotherapy (**D-F**) reveals new-onset enhancing T2/FLAIR hyperintense (white arrows) area anterior to the right frontal horn, involving genu of corpus callosum in the periresectional area, BT-RADS 3c score was assigned by RD2 as it was favouring tumour progression, even though the scan was done within 90 days of radiotherapy; while RD1 assigned 3a adhering to the timeline. Imaging (**G-J**) done 9 months after radiotherapy depicts further progression with new onset enhancing lesions (red arrows)

As a subset of our secondary objective, we calculated the predictive value of BT-RADS. Patients with score ≤ 3a exhibited an expected one-year overall survival probability of 94.8 (95% CI: 89.9-99.8); in contrast, those with scores ≥ 3b had a one-year OS probability of only 31.5 (95% CI: 4.0-59.0). This predictive value brings these management recommendations more in line with other structured reporting systems, like BIRADS or NIRADS.

We encountered certain issues during implementation of the scoring system. Firstly, we faced ambiguity when comparing immediate post-operative MRI as the baseline to any post-CRT (chemoradiotherapy) scan, in which there was resolution of haemorrhagic changes in the resection cavity, subdural/extradural haematoma, and decrease in oedema or perioperative ischaemic changes. These changes tended to settle down with time, as seen on the 6-week post-CRT scan, leading to an "apparent" improvement in imaging findings, even though there may not have been any change in the tumour burden per se. The guidelines in the BT-RADS standard scoring

template are unclear about improving imaging findings due to a decrease in the post-surgical findings, and hence the scores of either 2 or 1a were interchangeably assigned by each radiologist. Secondly, a potential limitation was the omission of advanced imaging techniques such as perfusion imaging and spectroscopy [12,13]. These techniques have been proven to offer additional insights as adjunctive tools to differentiate pseudoprogression versus true progression. Finally, time since radiotherapy (90 days) was the only parameter taken into consideration when assigning score 3a or 3b irrespective of the type of enhancement. This, in our opinion, was an oversimplification because a new unequivocal solid enhancing lesion must be given the benefit of doubt regarding disease progression (as demonstrated in our patient, Figure 3), even if it occurs within 90 days of radiation treatment. Our study has certain limitations. Firstly, the retrospective nature of the study is inherently susceptible to selection bias, and secondly, we had a limited sample size.

Conclusions

The BT-RADS structured reporting system was independently and externally validated to have good agreement between reporting radiologists. Despite the overall good agreement, variation rates escalated with worsening findings. The BT-RADS management recommendations for each score also showed near perfect concordance with decisions taken by our multidisciplinary team. There also seemed to be a potential predictive role in overall survival; however, additional data are required for validating the same.

Disclosure

The authors report no conflict of interest.

References

1. Claus EB, Walsh KM, Wiencke JK, et al. Survival and low-grade glioma: the emergence of genetic information. Neurosurg Focus 2015; 38: E6. doi: 10.3171/2014.10.FOCUS12367.

2. Crooms RC, Goldstein NE, Diamond EL, et al. Palliative care in high-grade glioma: a review. Brain Sci 2020; 10: 723. doi: 10.3390/brainsci10100723.

3. Smits M, van den Bent MJ. Imaging correlates of adult glioma genotypes. Radiology 2017; 284: 316-331.

4. Sahu A, Patnam NG, Goda JS, et al. Multiparametric magnetic resonance imaging correlates of isocitrate dehydrogenase mutation in WHO high-grade astrocytomas. J Pers Med 2022; 13: 72. doi: 10.3390/jpm13010072.

5. Kessler AT, Bhatt AA. Brain tumour post-treatment imaging and treatment-related complications. Insights into Imaging 2018; 9: 1057-1075.

6. Macdonald DR, Cascino TL, Schold SC Jr, et al. Response criteria for phase II studies of supratentorial malignant glioma. J Clin Oncol 1990; 8: 1277-1280.

7. Tensaouti F, Khalifa J, Lusque A, et al. Response assessment in neuro-oncology criteria, contrast enhancement and perfusion MRI for assessing progression in glioblastoma. Neuroradiology 2017; 59: 1013-1020.

8. Weinberg BD, Gore A, Shu HKG, et al. Management-based structured reporting of posttreatment glioma response with the brain tumor reporting and data system. J Am Coll Radiol 2018; 15: 767-771.

9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977: 159-174.

10. Cooper M, Hoch M, Abidi S, et al. NIMG-10. Brain tumor MRI structured reporting allows calculation of interrater agreement of patients reviewed at tumor board. Neuro Oncol 2019; 21 (Suppl 6): vi163. doi: 10.1093/neuonc/noz175.682.

11. Sahu A, Mathew R, Ashtekar R, et al. The complementary role of MRI and FET PET in high grade gliomas to differentiate recurrence from radionecrosis. Front Nucl Med 2023; 3: 1040998.

12. Gahramanov S, Muldoon LL, Varallyay CG, et al. Pseudoprogression of glioblastoma after chemo-and radiation therapy: diagnosis by using dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging with ferumoxytol versus gadoteridol and correlation with survival. Radiology 2013; 266: 842-852.

13. Elmogy SA, Mousa AE, Elashry MS, et al. MR spectroscopy in post-treatment follow up of brain tumors. Egypt J Radiol Nucl Med 2011; 42: 413-424.