HEPATOBILIARY-PANCREAS

LI-RADS treatment response categorization on gadoxetic acid-enhanced MRI: diagnostic performance compared to mRECIST and added value of ancillary features

Se Woo Kim [1,2] · Ijin Joo [1,2] · Hyo-Cheol Kim [1,2] · Su Joa Ahn [1,2] · Hyo-Jin Kang [1,2] · Sun Kyung Jeon [1,2] · Jeong Min Lee [1,2,3]

Abstract

Objectives This study was conducted in order to assess the performance of the Liver Imaging Reporting and Data System (LI-RADS) treatment response (TR) (LR-TR) categorization on gadoxetic acid-enhanced MRI (Gd-EOB-MRI) for detecting viable tumors in hepatocellular carcinoma (HCC) treated with locoregional treatment (LRT) and to investigate the added value of ancillary features (AFs) to conventional enhancement-based criteria.

Methods This retrospective study included 183 patients with Gd-EOB-MRI after LRT for HCC and appropriate reference standards for tumor viability (84 viable and 99 nonviable). Two independent radiologists assigned per-lesion mRECIST and TR categories (TR-nonviable, TR-equivocal, or TR-viable) according to the LR-TR algorithm and modified LR-TR algorithms including mLR-TR(TP) allowing transitional phase (TP) washout and mLR-TR(AF) allowing category adjustment by applying AFs. Diagnostic performances of imaging criteria were compared using the Cochran's Q test with post hoc analysis.

Results For detecting viable tumors, LR-TR-viable resulted in sensitivities of 64.5%/39.3% and specificities of 98.0%/98.0% in reviewers 1/2. In comparison to LR-TR-viable, mRECIST-viable, mLR-TR(TP)-viable, and mLR-TR(AF)-viable showed significantly higher sensitivities (92.9%/94.0%, 77.4%/56.6%, and 86.9%/83.3% in reviewers 1/2) (ps < 0.001). The specificity of mRECIST-viable (73.7%/62.6%) was significantly lower than that of LR-TR-viable (ps < 0.001), while those of mLR-TR(TP)-viable and mLR-TR(AF)-viable were greater than 95% (98.0%/96.0% and 97.0%/96.0%), statistically equivalent to LR-TR-viable (ps > 0.05). TR-equivocal was least assigned on mLR-TR(AF) (1.1%/7.7%) than LR-TR (15.8%/32.2%) or mLR-TR(TP) (6.6%/23.5%) in both reviewers.

Conclusion The LR-TR algorithm on Gd-EOB-MRI provides a specific diagnosis of viable tumor but with limited sensitivity. By applying AFs in the category adjustment, more sensitive and confident diagnosis can be achieved without significant loss of specificity.

Key Points

• *The LI-RADS treatment response (LR-TR) algorithm on Gd-EOB-MRI provides a highly specific diagnosis of viable HCC but with limited sensitivity.*

• *The inferior sensitivity of LR-TR-viable category to that of mRECIST can be improved by applying ancillary features in the category adjustment.*

Keywords Hepatocellular carcinoma · Magnetic resonance imaging · Liver · Practice guideline

---

✉ Ijin Joo
hijijin@gmail.com

1 Department of Radiology, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 03080, South Korea

2 Department of Radiology, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, South Korea

3 Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 03080, South Korea

Abbreviations

AP	Arterial phase
APHE	Arterial phase hyperenhancement
DWI	Diffusion-weighted imaging
Gd-EOB-MRI	Gadoxetic acid-enhanced magnetic resonance imaging
GRE	Gradient-echo
HASTE	Half-Fourier acquisition single-shot turbo spin-echo
HBP	Hepatobiliary phase
HCC	Hepatocellular carcinoma
LI-RADS	Liver Imaging Reporting and Data System
LRT	Locoregional treatment
LR-TR	LI-RADS treatment response
MDCT	Multidetector computed tomography
mRECIST	Modified response evaluation criteria in solid tumors
MRI	Magnetic resonance imaging
PVP	Portal venous phase
RFA	Radiofrequency ablation
TACE	Trans-catheter arterial chemoembolization
TP	Transitional phase

Introduction

Recently, locoregional therapy (LRT) including radiofrequency ablation (RFA) and trans-catheter arterial chemoembolization (TACE) has been increasingly used for the treatment of hepatocellular carcinoma (HCC), even in patients who are not indicated for surgical resection, although their indications are different according to guidelines [1–4]. Moreover, LRT can also be applied to recurrent tumors [5, 6] and as a bridging therapy to liver transplantation [2, 7]. After LRT, detection of residual viable tumor and local recurrence are critical for the timely retreatment of HCC, transplantation eligibility, and prediction of patients' prognosis after the subsequent hepatic resection or transplantation [8, 9]. At present, contrast-enhanced multidetector computed tomography (MDCT) and magnetic resonance imaging (MRI) are the standard imaging modalities used at follow-ups after LRT, which show good performances in detecting viable tumors [10, 11].

The Liver Imaging Reporting and Data System (LI-RADS) has recently been developed to standardize the interpretation of liver findings in patients at a high risk for HCC [12]. Since the 2017 version, a new algorithm of LI-RADS CT/MRI treatment response (LR-TR) has been introduced to guide the image interpretation of treated observations after LRT. The LR-TR algorithm adopts the concept of enhancement-based assessment, instead of size-based assessment, which is similar to the most widely used criteria for patients treated with LRT, the modified Response Evaluation Criteria in Solid Tumors

(mRECIST) [13]. However, the LR-TR algorithm is significantly different from mRECIST as (i) assessment is on a per-lesion basis, not on a per-patient basis; (ii) the observation can be categorized as "equivocal" if the radiologist is unsure about the tumor viability, which may improve the specificity of the viable category; and (iii) not only arterial phase hyperenhancement (APHE) but also "washout appearance" and "enhancement similar to pretreatment" are added as diagnostic features of the viable tumor, which may improve the sensitivity [12]. Given its unique features, the LR-TR algorithm, as a new diagnostic algorithm for treated observations after LRT, should be investigated for its diagnostic performance. A few recent studies revealed that the LR-TR algorithm provides good performances for histologic tumor viability in pretransplantation CT or MRI [14, 15]. However, considering that the LR-TR algorithm primarily aims to assess gross viable tumors rather than histological viability [12], its performance on the posttreatment surveillance setting not limited to pretransplant work-up needs to be determined.

Recently, gadoxetic acid-enhanced MRI (Gd-EOB-MRI) has been increasingly used, as it enables a highly sensitive diagnosis of HCC by providing the hepatobiliary phase (HBP). Moreover, multiparametric Gd-EOB-MRI includes dynamic phases, T2-weighted and T1-weighted dual gradient-echo (GRE) in-phase and out-of-phase imaging, and diffusion-weighted imaging (DWI) [16]. On Gd-EOB-MRI, viable tumors typically show APHE, hypointensity in the portal venous phase (PVP), hypointensity in the transitional phase (TP), hypointensity in the HBP, and restricted diffusion [17]. However, the current LR-TR algorithm only adopts enhancement patterns, whereas the LI-RADS algorithm for untreated observations allows the optional use of ancillary features (AFs) in the category adjustment [12, 18]. We postulated that in the treated observation, a combined interpretation of MRI enhancement features and AFs might help provide a more accurate diagnosis of viable tumors.

Therefore, this study aimed to assess the performance of the LR-TR algorithm on Gd-EOB-MRI for detecting viable tumors in HCC treated with LRT and to investigate the added value of AFs to conventional enhancement-based criteria.

Materials and methods

Patients

Our institutional review board approved this retrospective study, and the requirement for written informed consent was waived. The inclusion criteria in this study were (i) patients with available Gd-EOB-MRI after LRT for HCC and (ii) an appropriate reference standard for tumor viability (viable or nonviable) for the treated observation. In a retrospective search of our prospectively maintained database, we found

316 patients who had undergone initial LRT (TACE or RFA) for HCC in our institution between January 2015 and December 2016 for HCCs. Pre-LRT diagnosis of HCC was made by typical imaging findings (LR-5 according the CT/MRI LI-RADS v2018 [19] or "definite HCC" on CT or MRI according to Korean Liver Cancer Association and National Cancer Center guidelines v2018 [20]) or by histologic confirmation. The exclusion criteria were (i) patients with diffuse infiltrative tumor and (ii) MRI of poor image quality (Fig. 1).

Reference standard

Reference standards for "viable" tumors in treated observations included (i) histologic confirmation of viable HCC (< 6 weeks from the MRI) or (ii) presence of well-defined strong tumor stain on cone-beam CT angiography followed by TACE (< 6 weeks from the MRI) which also shows dense accumulation of iodized oil in the target lesion on post-TACE unenhanced CT (< 2 weeks from TACE) [21–23]. Digital subtraction angiography immediately before TACE was also reviewed to differentiate tumor stain from pre-existing iodized oil accumulation. On the other hand, reference standards for "nonviable" tumors included (i) histologic confirmation of total tumor necrosis (< 6 weeks from the MRI) or (ii) stable or decreased lesion size on the follow-up dynamic CT or MRI (> 6 months from the MRI) without any additional treatment. In patients with multiple sets of MRIs with reference standards for tumor viability, the first MRI containing at least one confirmed "viable" tumor was prioritized. When a patient had only "nonviable" lesions in the follow-ups, the first MRI after LRT was chosen. In addition, in patients with multiple lesions meeting the inclusion criteria, one lesion for each patient was selected, with priority to "viable" tumor, followed by the largest lesion

after LRT. The aforementioned selection process was performed by one radiologist (S.W.K.) who was unblinded to the patients' medical records and did not participate in the imaging review.

Finally, a total of 183 patients (133 men and 50 women; mean age 59.9 years; age range 27–86 years), each with one target-treated observation on Gd-EOB-MRI after LRT (84 viable and 99 nonviable), were included in this study. Table 1 shows the demographic and clinical characteristics of the study participants.

Image acquisition on Gd-EOB-MRI

Gd-EOB-MRI was performed using either a 3.0- or 1.5-T MR scanner: 3.0-T Magnetom Skyra, Magnetom Verio, Magnetom Trio Tim, and Biograph mMR (Siemens Healthineers; $n = 39$, 26, 10, and 10, respectively); 3.0-T Achieva CX and Ingenia (Philips Healthcare; $n = 42$ and 24, respectively); and 3.0-T Discovery MR 750w and 1.5-T Signa HDxT (GE Healthcare; $n = 22$ and 10, respectively). Our routine liver MRI protocol consisted of the following sequences: respiratory-triggered T2-weighted fast spin-echo and half-Fourier acquisition single-shot turbo spin-echo (HASTE) sequences, free-breathing DWI using b values of 0 and 800 s/mm^2, and breath-hold T1-weighted dual GRE in- and opposed-phase imaging. For dynamic imaging, patients received 0.1 mL/kg of intravenous gadoxetic acid (Primovist, Bayer Healthcare) at the rate of 1.0 mL/s, followed by 20 mL of saline flush. Pre- and postcontrast images using a T1-weighted, fat-suppressed, 3D GRE sequence were obtained. Arterial phase (AP) images were obtained 7 s after the contrast medium reached the distal thoracic aorta, as observed on real-time MRI fluoroscopic monitoring, and PVP, TP, and HBP images were obtained approximately 50–60 s, 3 min, and 20 min after the contrast medium injection, respectively.

Fig. 1 Flow chart of patient selection. HCC = hepatocellular carcinoma, LRT = locoregional treatment, Gd-EOB-MRI = gadoxetic acid-enhanced MRI
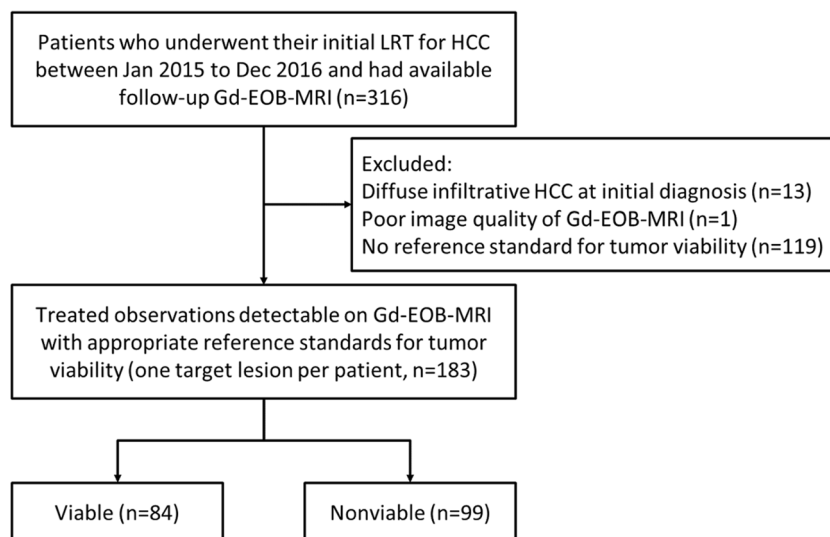
Table 1 Patient characteristics

Characteristics		Patients ($n = 183$)
Age (years), mean ± standard deviation		59.9 ± 10.8
Male:female		133:50
Child–Pugh classification	A	92.9 (170/183)
	B	7.1 (13/183)
Cause of chronic liver disease	Chronic hepatitis B	60.7 (111/183)
	Chronic hepatitis C	13.7 (25/183)
	Alcoholic liver disease	18.6 (34/183)
	Others	7.1 (13/183)
Reference standards for tumor viability		
Viable ($n = 84$)	Histopathology	14.3 (12/84)
	Cone-beam CT angiography	85.7 (72/84)
Nonviable ($n = 99$)	Histopathology	9.1 (9/99)
	Follow-up imaging (> 6 months)	90.9 (90/99)
Previous locoregional treatment	Conventional TACE	72.1 (132/183)
	DEB-TACE	2.7 (5/183)
	RFA	23.0 (42/183)
	TACE+RFA	2.2 (4/183)

Note: data in parentheses are numbers used to calculated percentages

TACE = trans-catheter arterial embolization, DEB = drug-eluting beads, RFA = radiofrequency ablation

Acquisition of 3D GRE images for each dynamic phase and for HBP was completed during a single breath-hold (15–20 s) at the end of expiration.

Image interpretation

Two reviewers (S.K.J. and H.J.K., with 5 and 7 years of experience, respectively, in liver MRI) independently evaluated MRI features and determined the TR category for each observation after LRT. Before the imaging review, reviewers were trained to understand the principles of LR-TR algorithm and were provided with an interactive lecture on eight practical cases selected from a different period in our study population. For analysis, reviewers were provided with captured HBP images of annotated target lesions and information of the type of LRT (TACE or RFA) for each lesion. However, they were blinded to the presence and absence of viable tumors in the target lesions. Assessed MRI features included the nodular, mass-like, or thick irregular tissue in or along the treated lesion with APHE, washout appearance on PVP, hypointensity on TP, hypointensity on HBP, restricted diffusion, and intermediate hyperintensity on T2-weighted imaging [12, 18]. If APHE was assessed to be present in or around the lesion, it was considered mRECIST-viable, otherwise mRECIST-nonviable. Of note, while mRECIST primarily aims to determine overall disease status per patient [13], this study only adopted the mRECIST principle for per-lesion basis interpretation of tumor viability [15].

For each observation, the reviewers were asked to assign the TR category (TR-nonviable, TR-equivocal, or TR-viable)

according to three different criteria: (i) LR-TR, (ii) modified LR-TR with TP washout (mLR-TR(TP)), and (iii) modified LR-TR with AFs (mLR-TR(AF)) (Figs. 2 and 3). To be more specific, TR-viable on the LR-TR algorithm indicated nodular, mass-like, or thick irregular tissue showing either APHE or washout, in or along the treated lesion. Following the LI-RADS principle for the use of Gd-EOB-MRI, washout appearance was determined only on PVP [12]. In this study, previous imaging studies were not provided to the reviewers, and therefore, "enhancement similar to pretreatment," which is one of the imaging criteria for the current LR-TR algorithm [12], could not be assessed. The mLR-TR(TP) criteria used an extended definition of washout on Gd-EOB-MRI, which adopted "TP hypointensity" as an equivalent to the "washout appearance" on PVP based on the previous studies showing improved sensitivity for HCC diagnosis in untreated observations [24, 25]. Thus, mLR-TR(TP) criteria for viable category included APHE or washout appearance on PVP or TP hypointensity. In the mLR-TR(AF) algorithm, AFs favoring malignancy in general (T2-weighted intermediate hyperintensity, presence of restricted diffusion in high b-value DWI, and HBP hypointensity) and an AF favoring benignity (isointensity on HBP) were selectively adopted from the LI-RADS v2018 for untreated observations [12, 18] and were optionally applied in the category adjustment (upgrade or downgrade between nonviable and equivocal, equivocal and viable, and nonviable and viable). Within each algorithm, if a radiologist is unsure between categories, TR-equivocal was chosen following the tie-breaking rule of LR-TR suggested by the LI-RADS [12].

Fig. 2 Hepatocellular carcinoma treated with trans-catheter arterial chemoembolization (TACE) in an 80-year-old male patient. On Gd-EOB-MRI, a TACE-treated lesion in segment IV shows hyperintensity (asterisk) on the precontrast T1-weighted image (**a**). Along the anteromedial margin of the treated lesion, there is a nodular tissue (arrow) with arterial phase hyperenhancement (APHE) (**b**), but without definite washout appearance on the portal venous phase (PVP) (**c**). On the transitional phase (TP) (**d**) and hepatobiliary phase (HBP) (**e**), it shows hypointensity (arrow). This treated observation is assigned as mRECIST-viable and LR-TR-viable based on the finding of APHE. When applying the modified LR-TR criteria, it is assigned as the viable category using mLR-TR(TP) criteria which allows TP to determine washout appearance as well as using mLR-TR(AF) criteria which allows the application of ancillary features in the category adjustment. Cone-beam CT hepatic angiography (**f**) image demonstrates the presence of tumor stain (arrow), and a subsequent TACE was performed for that tumor stain. Post-TACE unenhanced CT revealed the dense accumulation of iodized oil in the tumor (not shown)
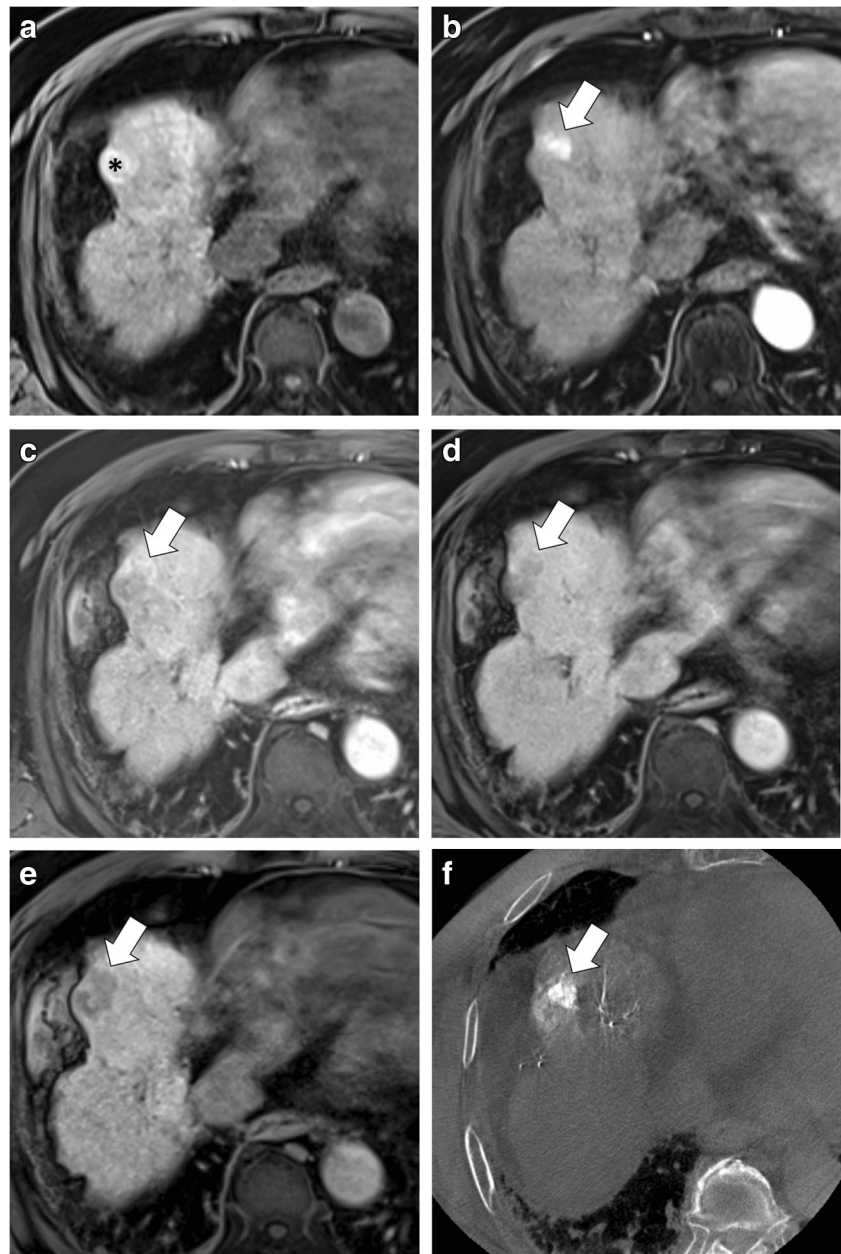
Statistical analysis

Sensitivity and specificity of the TR-viable category based on LR-TR criteria and those based on the modified criteria for detecting viable tumors were compared to each other and to those based on mRECIST (i.e., presence of APHE [13]), using the Cochran's Q test followed by post hoc pairwise McNemar's test. Additionally, a subgroup analysis was performed according to the LRT methods (RFA group: received only RFA, $n = 42$; TACE group: received TACE, with or without RFA, $n = 141$) to compare sensitivity and specificity using the chi-square test. For each MRI feature, sensitivity and specificity for the diagnosis of viable tumor were calculated. To achieve a

consensus on the interpretation of MRI features, the discrepancies were resolved by a third reviewer (S.J.A. with 10 years of experience in liver MRI). Weighted and unweighted kappa statistics were used to evaluate the interobserver agreement for TR categorization (0: TR-nonviable, 1: TR-equivocal, and 2: TR-viable) and for the assessment of MRI features (0: absent and 1: present). The level of agreement was interpreted as poor ($\kappa \leq 0.2$), fair ($0.2 < \kappa \leq 0.4$), moderate ($0.4 < \kappa \leq 0.6$), good ($0.6 < \kappa \leq 0.8$), and excellent ($0.8 < \kappa \leq 1.0$). Two commercially available software packages (SPSS, Version 21.0 for Windows, SPSS; MedCalc Statistical Software, Version 17.6, MedCalc Software) were used for the statistical analysis. A p value less than 0.05 was considered statistically significant,

Fig. 3 Hepatocellular carcinoma treated with radiofrequency ablation (RFA) in a 66-year-old female patient. On Gd-EOB-MRI, a small vague arterially enhancing area (arrow) is observed along the posteromedial margin of the ablation defect (asterisk) (**a**), which does not show washout appearance on the PVP (**b**). This observation is categorized as mRECIST-viable, but LR-TR-equivocal as its enhancement is atypical for treatment-specific change or viable tumor. The arterially enhancing area shows ancillary features favoring malignancy in general including hypointensity on the HBP (**c**), intermediate hyperintensity on T2-weighted image (**d**), and restricted diffusion ($b = 800$ s/mm^2) (**e**). Therefore, when using mLR-TR(AF), the viable category can be assigned to this observation. The presence of a viable tumor (arrow) is confirmed by the subsequent cone-beam CT hepatic angiography (**f**) and post-TACE unenhanced CT (dense accumulation of iodized oil) (not shown)

except for the post hoc McNemar's test (p value less than 0.013 [0.05/4] based on Bonferroni correction).

Result

Diagnostic performances of the LR-TR algorithm on Gd-EOB-MRI

According to the LR-TR criteria, viable tumors were categorized as viable, equivocal, and nonviable in 66.7% (56/84), 21.4% (18/84), and 11.9% (10/84) by reviewer 1 and 40.5% (34/84), 52.4% (44/84), and 7.1% (6/84) by reviewer 2, respectively. Further, nonviable tumors were categorized as

viable, equivocal, and nonviable in 2.0% (2/99), 11.1% (11/99), and 86.9% (86/99) by reviewer 1 and 2.0% (2/99), 15.2% (15/99), and 82.8% (82/99) by reviewer 2, respectively. Accordingly, the LR-TR-viable category resulted in the sensitivities of 66.7% (56/84) and 40.5% (34/84), specificities of 98.0% (97/99) and 98.0% (97/99), and accuracies of 83.6% (153/183) and 71.6% (131/183) in reviewers 1 and 2, respectively (Table 2).

Comparison of performances of the LR-TR algorithm and mRECIST

The mRECIST criteria for viable tumors (presence of APHE) showed sensitivities of 92.9% (78/84) and 94.0% (79/84) and

Table 2 Comparison of diagnostic performance of treatment response-viable category according to different Gd-EOB-MRI criteria

	Reviewer 1			Reviewer 2		
	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
Imaging criteria						
(i) LR-TR-viable	66.7 (56/84)	98.0 (97/99)	83.6 (153/183)	40.5 (34/84)	98.0 (97/99)	71.6 (131/183)
(ii) mLR-TR(TP)-viable	77.4 (65/84)	98.0 (97/99)	88.5 (162/183)	56.6 (47/84)	96.0 (95/99)	77.6 (142/183)
(iii) mLR-TR(AF)-viable	86.9 (73/84)	97.0 (96/99)	92.3 (169/183)	83.3 (70/84)	96.0 (95/99)	90.2 (165/183)
(iv) mRECIST-viable	92.9 (78/84)	73.7 (73/99)	82.5 (151/183)	94.0 (79/84)	62.6 (62/99)	77.0 (141/183)
p values						
Cochran's *Q* test	*< 0.001*	*< 0.001*	*0.001*	*< 0.001*	*< 0.001*	*< 0.001*
Post hoc (McNemar's test)						
(i) versus (ii)	*0.004*	> 0.999	*0.012*	*< 0.001*	0.500	*0.007*
(i) versus (iii)	*< 0.001*	> 0.999	*< 0.001*	*< 0.001*	0.500	*< 0.001*
(ii) versus (iii)	*0.008*	> 0.999	0.092	*< 0.001*	> 0.999	*< 0.001*
(i) versus (iv)	*< 0.001*	*< 0.001*	0.885	*< 0.001*	*< 0.001*	0.314
(ii) versus (iv)	*< 0.001*	*< 0.001*	0.099	*< 0.001*	*< 0.001*	> 0.999
(iii) versus (iv)	0.063	*< 0.001*	*0.001*	*0.004*	*< 0.001*	*< 0.001*

Note: data in parentheses are numbers used to calculate percentages. Numbers in italics indicate statistical significance

LR-TR = Liver Imaging Reporting and Data System (LI-RADS) treatment response, *mLR-TR* = modified LR-TR, *TP* = transitional phase, *AF* = ancillary feature, *mRECIST* = modified response evaluation criteria in solid tumors

specificities of 73.7% (73/99) and 62.6% (62/99) in reviewers 1 and 2, respectively (Table 2). When comparing mRECIST and LR-TR, sensitivities were significantly higher in mRECIST-viable (*p*s < 0.001), while specificities were significantly higher in LR-TR-viable in both reviewers (*p*s < 0.001). The mRECIST-viable lesions (*n* = 104 and 116 in reviewers 1 and 2, respectively) were assigned as LR-TR-viable in 54.8% (57/104) and 31.0% (36/116), LR-TR-equivocal in 25.0% (26/104) and 49.1% (57/116), or LR-TR-nonviable in 20.2% (21/104) and 19.8% (23/116) by reviewers 1 and 2, respectively.

Comparison of diagnostic performances of different TR criteria

Table 2 shows the diagnostic performances of the viable categories of different TR criteria and their comparisons. mLR-TR(TP)-viable showed significantly higher sensitivities and accuracies than LR-TR-viable in both reviewers (*p*s < 0.05). In addition, mLR-TR(AF)-viable showed significantly higher sensitivities and accuracies than mLR-TR(TP)-viable and LR-TR-viable in both reviewers (*p*s < 0.05). However, the viable categories of LR-TR, mLR-TR(TP), and mLR-TR(AF) showed comparable specificities, greater than 95%, in both reviewers (*p*s > 0.05). In addition, sensitivities and specificities of the viable category were not significantly different between different treatment methods (RFA only versus

TACE with or without RFA) in either of LR-TR, mLR-TR(TP), and mLR-TR(AF) (Table E1).

In the assignment of category as viable, equivocal, or nonviable on different TR criteria, proportions of the equivocal category decreased in the order of LR-TR, mLR-TR(TP), and mLR-TR(AF) in both reviewers (15.8% [29/183], 6.6% [12/183], and 1.1% [2/183], respectively, in reviewer 1; 32.2% [59/183], 23.5% [43/183], and 7.7% [14/183], respectively, in reviewer 2) (Table E2). When comparing the LR-TR category and the mLR-TR(AF) category, category adjustments by applying AFs were mostly toward either from equivocal to viable or from equivocal to nonviable (Table E3). The interobserver agreements for the assignment of category as viable, equivocal, or nonviable were moderate in LR-TR, good in mLR-TR(TP), and good in mLR-TR(AF) (Table E2).

Diagnostic performance of each Gd-EOB-MRI feature for predicting tumor viability

Diagnostic performances of the Gd-EOB-MRI features assessed for viable tumors are described in Table 3. Among all the imaging features assessed herein, APHE showed the highest sensitivity (96.4%, 81/84), whereas washout appearance in the PVP showed the highest specificity (92.9%, 92/99). Compared to PVP washout, TP hypointensity showed significantly higher sensitivity (82.1% [69/84] versus 67.9% [57/84], *p* < 0.001) and the same specificity (92.9% [92/99]

Table 3 Diagnostic performance and interobserver agreement of Gd-EOB-MRI features

Imaging features (consensus interpretation)	Diagnostic performance for viable tumors			Interobserver agreement
	Sensitivity (%)	Specificity (%)	Accuracy (%)	κ (95% CI)
APHE	96.4 (81/84)	68.7 (68/99)	81.4 (149/183)	0.61 (0.50–0.73)
PVP washout	67.9 (57/84)	92.9 (92/99)	81.4 (149/183)	0.56 (0.43–0.69)
TP hypointensity	82.1 (69/84)	92.9 (92/99)	88.0 (161/183)	0.56 (0.44–0.68)
HBP hypointensity	95.2 (80/84)	81.8 (81/99)	88.0 (161/183)	0.70 (0.59–0.80)
Intermediate hyperintensity on T2WI	84.5 (71/84)	83.8 (83/99)	84.2 (154/183)	0.53 (0.41–0.66)
Restricted diffusion	83.3 (70/84)	91.9 (91/99)	88.0 (161/183)	0.67 (0.55–0.78)

Note: data in parentheses are numbers used to calculate percentages

APHE = arterial phase hyperenhancement, *PVP* = portal venous phase, *TP* = transitional phase, *HBP* = hepatobiliary phase, *T2WI* = T2-weighted imaging, *CI* = confidence interval

versus 92.9% [92/99], $p > 0.999$). In addition, HBP hypointensity showed higher sensitivity and lower specificity compared to TP hypointensity or washout appearance in the PVP (Table 3). Interobserver agreements were moderate in PVP washout, TP hypointensity, and intermediate hyperintensity on T2-weighted image and good in APHE, HBP hypointensity, and restricted diffusion (Table 3).

Discussion

This retrospective study validated the performance of the LR-TR algorithm, recently introduced imaging criteria for tumor response of HCC after LRT, using Gd-EOB-MRI. According to our study results, the LR-TR-viable category on Gd-EOB-MRI, i.e., APHE or PVP washout, showed very high specificity but insufficient sensitivity in the diagnosis of viable tumors. In the assignment of the category as viable, equivocal, or nonviable, the LR-TR criteria demonstrated moderate interobserver agreement and resulted in a substantial proportion of observations being assigned under the equivocal category. In addition, this study also evaluated other Gd-EOB-MRI response criteria modified from the LR-TR: mLR-TR(TP) (allowing both PVP and TP to determine washout) and mLR-TR(AF) (applying AFs for category adjustment). Compared to the LR-TR criteria, the viable category of the modified criteria significantly improved sensitivity, which was the highest in mLR-TR(AF), without significant loss of specificity.

In this study, the viable category of the LR-TR criteria showed a sensitivity of 40.5–66.7% for detecting viable tumors. Given that the LR-TR-viable category was determined by either the presence of APHE or washout appearance on the PVP of Gd-EOB-MRI, the low sensitivity may be due to the limitation of either AP or PVP to demonstrate the typical enhancement features of viable HCCs. However, as for APHE, our study demonstrated that the mRECIST criteria

(presence of APHE) resulted in very high sensitivity (92.9–94.0%). Thus, the lower sensitivity of LR-TR-viable as compared to mRECIST in our study indicates that although the reviewers detected APHE, they did not assign them as LR-TR-viable in some observations. Indeed, unlike mRECIST, the LR-TR algorithm considers the radiologists' certainty for tumor viability which may increase the specificity of LR-TR-viable by reducing false-positive diagnosis in treatment-related changes such as perilesional hyperemia [17]. However, at the same time, it may result in the assignment of the equivocal or nonviable category for observations with APHE but not characteristic for viable tumors, leading to reduction in the sensitivity of the viable category as shown in our study results. On the other hand, the viable category on mLR-TR(TP), which have adopted TP hypointensity as an equivalent to PVP washout, increased the sensitivity in comparison to that of LR-TR. This result was quite the same as expected because the only difference between LR-TR and mLR-TR(TP) is the allowed phases to determine washout appearance ("PVP only" versus "PVP or TP"), and was in good agreement with previous reports on untreated observations [24, 25]. Moreover, the sensitivity was even higher in mLR-TR(AF) than in mLR-TR(TP) in our study results, which suggests the role of AFs for detecting viable tumors on T2-weighted imaging, HBP, and DWI, as shown in previous studies [26–28].

The LR-TR-viable category resulted in a high specificity of 98.0% in this study, which was significantly higher than that of the mRECIST criteria (62.6–73.7%). In addition, the high specificity of the LR-TR criteria was maintained in mLR-TR(TP) and mLR-TR(AF), whereas the sensitivity was significantly increased in these modified criteria. In clinical practice, a specific diagnosis of viable tumors in treated observations is important to avoid unnecessary treatment. According to the LI-RADS for untreated observations, when using Gd-EOB-MRI, washout appearance is recommended to be evaluated only in the PVP and ancillary features cannot be used to

upgrade to LI-RADS 5 (definitely HCC), which primarily aims to avoid false-positive diagnosis of HCC in cases that mimic HCC [29]. However, as opposed to the CT/MRI LI-RADS for untreated observations focusing on differentiating HCC from benign lesions or non-HCC malignancies, the LR-TR-viable category focuses on the presence or absence of viable tumors in treated HCCs. Therefore, in the LR-TR algorithm, high specificity can be achieved by avoiding false-positive diagnosis of benign treatment-related changes, while the differential diagnosis between HCC and non-HCC malignancy is not a main concern in the LR-TR algorithm. Indeed, previous studies reported the usefulness of AFs of DWI or HBP for differentiating between arterially enhancing pseudolesions and HCCs [30, 31]. This supports the use of AFs favoring malignancy in general or favoring benignity in treated observations, which can provide high sensitivity and high specificity, as shown in our study. Notably, specificities of our TR criteria were superior to those reported in recent MRI studies using imaging criteria other than the LR-TR algorithm (specificity, approximately 75–85%) [28, 32]. This difference might be caused by the presence of equivocal category in LR-TR and by different reference standards used for determining tumor viability. The aforementioned previous studies [28, 32] used only explant histopathology as reference standards, whereas our study included both histological diagnosis and clinical imaging-based diagnosis. Therefore, our study results may not be directly applied to predict pathologic complete response. However, considering that the LR-TR algorithm essentially aims to assess gross viable tumors, not histological viability [33], we believe that our study population may be more appropriate for the validation of the LR-TR criteria in the general setting of posttreatment surveillance.

In the present study, proportions of the equivocal category decreased in the order of LR-TR, mLR-TR(TP), and mLR-TR(AF) criteria. Given that the equivocal category indicates that the radiologist is not confident whether the observation is viable or nonviable, our study results suggest that the mLR-TR(AF) algorithm may enable a more confident diagnosis for tumor viability by comprehensive interpretation of multiparametric Gd-EOB-MRI. In clinical practice, assignment of the equivocal category might pose the need for alternative or short-term follow-up imaging study, leading to delayed management decision. Therefore, for timely and confident management decision, mLR-TR(AF), which can reduce the frequency of equivocal category assignment without loss of diagnostic accuracy, is of great clinical value. Moreover, our study showed that interobserver agreement for category assignment was good in mLR-TR(AF), whereas it was moderate in LR-TR. Given that interobserver reliability is critical for a consistent interpretation of tumor response, mLR-TR(AF) may be more appropriate in multicenter studies for patient allocation or outcome analysis; however, it should be further validated.

Our study has several limitations. First, this retrospective study had partial verification bias, as the diagnostic performance was assessed only in those with reference standards for tumor viability. Second, while we only included treated observations by RFA or TACE, i.e., more conventional LRTs, the LR-TR algorithm can also be applied in other LRTs such as microwave ablation, radioembolization, and external beam radiation therapy. Therefore, further studies are needed for those newer treatment methods. Third, as mentioned in the "Materials and methods," this study did not include pretreatment images and, therefore, lacks the validation of the LR-TR-viable criteria regarding "enhancement similar to pretreatment." Fourth, we only assessed Gd-EOB-MRI and lacked comparison with other imaging modalities such as dynamic contrast-enhanced CT and MRI with extracellular contrast agents. Intraindividual comparison studies are required to compare the diagnostic performance of LR-TR according to imaging modalities.

In conclusion, the LR-TR algorithm showed high specificity but limited sensitivity for detection of viable tumors in Gd-EOB-MRI. The addition of ancillary features favoring malignancy in general or favoring benignity in the category adjustment resulted in a more sensitive and confident diagnosis of viable tumors, without significant loss of specificity.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Ijin Joo.

Conflict of interest The authors declare that they have no conflict of interest.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology
• Retrospective
• Diagnostic or prognostic study
• Performed at one institution

References

1. Korean Liver Cancer Study Group (KLCSG); National Cancer Center, Korea (NCC) (2015) 2014 Korean Liver Cancer Study Group-National Cancer Center Korea practice guideline for the

management of hepatocellular carcinoma. Korean J Radiol 16:465–522

2. Heimbach JK, Kulik LM, Finn RS et al (2018) AASLD guidelines for the treatment of hepatocellular carcinoma. Hepatology 67:358–380

3. Yu SJ (2016) A concise review of updated guidelines regarding the management of hepatocellular carcinoma around the world: 2010-2016. Clin Mol Hepatol 22:7

4. Bruix J, Sherman M (2011) Management of hepatocellular carcinoma: an update. Hepatology 53:1020–1022

5. Chen R, Gan Y, Ge N et al (2016) Transarterial chemoembolization versus radiofrequency ablation for recurrent hepatocellular carcinoma after resection within Barcelona Clinic liver cancer stage 0/a: a retrospective comparative study. J Vasc Interv Radiol 27:1829–1836

6. Koh PS, Chan AC, Cheung TT et al (2015) Efficacy of radiofrequency ablation compared with transarterial chemoembolization for the treatment of recurrent hepatocellular carcinoma: a comparative survival analysis. HPB (Oxford) 18(1):72–78

7. Lee MW, Raman SS, Asvadi NH et al (2017) Radiofrequency ablation of hepatocellular carcinoma as bridge therapy to liver transplantation: a 10-year intention-to-treat analysis. Hepatology 65:1979–1990

8. Ho M-H, Yu C-Y, Chung K-P et al (2011) Locoregional therapy-induced tumor necrosis as a predictor of recurrence after liver transplant in patients with hepatocellular carcinoma. Ann Surg Oncol 18:3632–3639

9. Allard M-A, Sebagh M, Ruiz A et al (2015) Does pathological response after transarterial chemoembolization for hepatocellular carcinoma in cirrhotic patients with cirrhosis predict outcome after liver resection or transplantation? J Hepatol 63:83–92

10. Yu JS, Kim JH, Chung JJ, Kim KW (2009) Added value of diffusion-weighted imaging in the MRI assessment of perilesional tumor recurrence after chemoembolization of hepatocellular carcinomas. J Magn Reson Imaging 30:153–160

11. Kloeckner R, Otto G, Biesterfeld S, Oberholzer K, Dueber C, Pitton MB (2010) MDCT versus MRI assessment of tumor response after transarterial chemoembolization for the treatment of hepatocellular carcinoma. Cardiovasc Intervent Radiol 33:532–540

12. Tang A, Bashir MR, Corwin MT et al (2017) Evidence supporting LI-RADS major features for CT-and MR imaging-based diagnosis of hepatocellular carcinoma: a systematic review. Radiology 286:29–48

13. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. Semin Liver Dis 30:52–60

14. Shropshire EL, Chaudhry M, Miller CM et al (2019) LI-RADS treatment response algorithm: performance and diagnostic accuracy. Radiology 292(1):226–234. https://doi.org/10.1148/radiol.2019182135:182135

15. Seo N, Kim MS, Park MS et al (2019) Evaluation of treatment response in hepatocellular carcinoma in the explanted liver with Liver Imaging Reporting and Data System version 2017. Eur Radiol. https://doi.org/10.1007/s00330-019-06376-5

16. Joo I, Lee JM (2016) Recent advances in the imaging diagnosis of hepatocellular carcinoma: value of gadoxetic acid-enhanced MRI. Liver Cancer 5:67–87

17. Yaghmai V, Besa C, Kim E, Gatlin JL, Siddiqui NA, Taouli B (2013) Imaging assessment of hepatocellular carcinoma response to locoregional and systemic therapy. AJR Am J Roentgenol 201:80–96

18. Cerny M, Chernyak V, Olivié D et al (2018) LI-RADS version 2018 ancillary features at MRI. Radiographics 38:1973–2001

19. Chernyak V, Fowler KJ, Kamaya A et al (2018) Liver Imaging Reporting and Data System (LI-RADS) version 2018: imaging of hepatocellular carcinoma in at-risk patients. Radiology 289:816–830

20. Korean Liver Cancer Association; National Cancer Center (2019) 2018 Korean Liver Cancer Association–National Cancer Center Korea practice guidelines for the management of hepatocellular carcinoma. Gut Liver 13:227

21. Bhattacharya S, Novell JR, Winslet MC, Hobbs KE (1994) Iodized oil in the treatment of hepatocellular carcinoma. Br J Surg 81:1563–1571

22. Yumoto Y, Jinno K, Tokuyama K et al (1985) Hepatocellular carcinoma detected by iodized oil. Radiology 154:19–24

23. Novell R, Dusheiko G, Hilson A, Dick R, Begent R, Hobbs K (1991) Lipiodol computed tomography for small hepatocellular carcinomas. Lancet 337:729

24. Min JH, Kim JM, Kim YK et al (2018) Prospective intraindividual comparison of magnetic resonance imaging with gadoxetic acid and extracellular contrast for diagnosis of hepatocellular carcinomas using the Liver Imaging Reporting and Data System. Hepatology 68:2254–2266

25. Joo I, Lee JM, Lee DH, Jeon JH, Han JK, Choi BI (2015) Noninvasive diagnosis of hepatocellular carcinoma on gadoxetic acid-enhanced MRI: can hypointensity on the hepatobiliary phase be used as an alternative to washout? Eur Radiol 25:2859–2868

26. Miyayama S, Yamashiro M, Nagai K et al (2016) Evaluation of tumor recurrence after superselective conventional transcatheter arterial chemoembolization for hepatocellular carcinoma: comparison of computed tomography and gadoxetate disodium-enhanced magnetic resonance imaging. Hepatol Res 46:890–898

27. Imai Y, Katayama K, Hori M et al (2017) Prospective comparison of Gd-EOB-DTPA-enhanced MRI with dynamic CT for detecting recurrence of HCC after radiofrequency ablation. Liver Cancer 6:349–359

28. Rimola J, Davenport MS, Liu PS et al (2018) Diagnostic accuracy of MRI with extracellular vs. hepatobiliary contrast material for detection of residual hepatocellular carcinoma after locoregional treatment. Abdom Radiol (NY) 44(2):549–558

29. Tang A, Bashir MR, Corwin MT et al (2018) Evidence supporting LI-RADS major features for CT- and MR imaging-based diagnosis of hepatocellular carcinoma: a systematic review. Radiology 286:29–48

30. Motosugi U, Ichikawa T, Sou H et al (2010) Distinguishing hypervascular pseudolesions of the liver from hypervascular hepatocellular carcinomas with gadoxetic acid-enhanced MR imaging. Radiology 256:151–158

31. Sun HY, Lee JM, Shin CI et al (2010) Gadoxetic acid-enhanced magnetic resonance imaging for differentiating small hepatocellular carcinomas (≤ 2 cm in diameter) from arterial enhancing pseudolesions: special emphasis on hepatobiliary phase imaging. Invest Radiol 45:96–103

32. Gordic S, Corcuera-Solano I, Stueck A et al (2017) Evaluation of HCC response to locoregional therapy: validation of MRI-based response criteria versus explant pathology. J Hepatol 67:1213–1221

33. Kulik L, Heimbach JK, Zaiem F et al (2018) Therapies for patients with hepatocellular carcinoma awaiting liver transplantation: a systematic review and meta-analysis. Hepatology 67:381–400