The LI-RADS v2018 Treatment Response algorithm is accurate for evaluating ablated hepatocellular carcinoma

Journal:	*Radiology*
Manuscript ID	RAD-19-1581.R1
Manuscript Type:	Original Research
Manuscript Categorization Terms:	Adults < 1. SUBJECT MATTER, MR-Imaging < 2. MODALITIES/TECHNIQUES, Abdomen/GI < 4. AREAS/SYSTEMS, Liver < 5. STRUCTURES, Tumor Response < 6. TOPICS

The LI-RADS v2018 MRI Treatment Response algorithm: evaluation of ablated hepatocellular carcinoma

Article type: Original research

Summary statement: **The Liver Imaging Reporting and Data System (LI-RADS) v2018 Treatment Response algorithm performs well for lesions treated by ablation when observations are assessed on MRI as Viable or Nonviable disease.**

Key Results:

1. After ablation, the sensitivity of the Liver Imaging Reporting and Data System (LI-RADS) Treatment Response Viable category for predicting residual tumor on MRI is 40-77%.
2. After ablation, the specificity of the Treatment Response Nonviable category for predicting complete tumor necrosis is 81%-85%.
3. After ablation, most Treatment Response Equivocal lesions (n=5/6) were incompletely necrotic on histopathology.

Abbreviations:

APHE – Arterial phase hyperenhancement

HCC – Hepatocellular carcinoma

LI-RADS – Liver Imaging Reporting and Data System

LR-TR – LI-RADS Treatment Response

TRA – Treatment Response Algorithm

Abstract

Background: The Liver Imaging Reporting and Data System (LI-RADS) Treatment Response Algorithm (TRA) is used to assess presumed hepatocellular carcinoma (HCC) after locoregional therapy, but its performance has not been extensively assessed.

Purpose: To assess the performance of the LI-RADS v2018 Treatment Response algorithm for evaluating hepatocellular carcinoma following ablation.

Materials and Methods: In this retrospective study, patients who underwent ablation therapy for presumed hepatocellular carcinoma followed by liver transplantation between January 2011 and December 2015 at a single tertiary care center were identified. Lesions were categorized as completely (100%) or incompletely (≤99%) necrotic based on transplant histology. Three radiologists assessed pre- and post-treatment MRI using LI-RADS v2018 and the TRA, respectively. Inter-reader agreement was assessed using Fleiss' Kappa. Performance characteristics for predicting necrosis category based on LI-RADS Treatment Response (LR-TR) category (Viable or Nonviable) were calculated using generalized mixed effects models to account for clustering by subject.

Results: 36 patients (mean age 58±5 years, 32 men) with 53 lesions were included. Inter-reader agreement for pre-treatment LI-RADS category was k=0.40 (95% confidence interval [CI] 0.15, 0.67; p<0.01) and was lower than the inter-reader agreement for TRA category (k=0.71 [95% CI: 0.59, 0.84]; p<0.01). After accounting for clustering by subject, the sensitivity of tumor necrosis across readers ranged from 40%-77% and the specificity ranged from 85%-97% when LR-TR Equivocal assessments were treated as Nonviable. When LR-TR Equivocal assessments were treated as Viable, the sensitivity of tumor necrosis across

readers ranged from 81%-87% and the specificity ranged from 81%-85% across readers. 11% (6/53) of treated lesions were LR-TR Equivocal by consensus, with most (5/6) incompletely necrotic on histopathology.

Conclusion: The Liver Imaging Reporting and Data System **(**LI-RADS) Treatment Response Algorithm predicts viable or nonviable hepatocellular carcinoma after ablation. Most ablated lesions rated as Treatment Response Equivocal were incompletely necrotic on histopathology.

Introduction

Hepatocellular carcinoma (HCC) is the third leading cause of cancer-related death worldwide [1, 2]. Due to its high morbidity, mortality, and progressive incidence, HCC represents an ongoing challenge for healthcare systems and has required continual expansion of treatment options [3]. Therapies such as transcatheter embolization, ablation, and external beam radiotherapy are frequently used for both definitive treatment of early stage HCC and neo-adjuvant therapy to downsize and/or control tumor burden prior to transplantation[4]. Given this increased application of locoregional therapies, understanding the appearance of treated lesions and response to therapy has become crucial in making management decisions. Existing guidelines such as the Response Evaluation Criteria In Solid Tumors and modified Response Evaluation Criteria In Solid Tumors emphasize patient-level response but do not address the variable appearances of individual lesions following treatment [5]. The growing use of locoregional therapy and difficulties in assessing response to treatment has led to the development of a Treatment Response Algorithm (TRA) in the Liver Imaging Reporting and Data System (LI-RADS) version 2018 (v2018), which is designed specifically to assist radiologists in interpretation of imaging of treated HCC.

LI-RADS provides interpretation and reporting guidelines for CT and MRI liver examinations performed on patients at risk of developing hepatocellular carcinoma (HCC) [6, 7]. It has been integrated into the most recent American Association for the Study of Liver Diseases HCC clinical practice guidance [8]. In LI-RADS v2018, the TRA takes into consideration that patients with HCC often undergo locoregional therapy and offers a comprehensive approach to standardize the assessment of treatment response of individual lesions using either contrast-enhanced CT or MRI. Using imaging features applied through a precise algorithm, LIRADS TRA allows categorization of treated lesions as Viable, Nonviable, or Equivocal [3]. However, despite the adoption of LI-RADS TRA at many

institutions, prediction of histological response based on the TRA has not been assessed extensively. The evaluation of treated lesions via thorough examination of explanted livers is critical for validation of the treatment response algorithm.

The purpose of this study is to assess the performance of the LI-RADS v2018 Treatment Response algorithm for evaluating hepatocellular carcinoma treated with ablation therapy.

Materials and Methods

Study Design

This was an Institutional Review Board (IRB) approved and HIPAA compliant retrospective study. The requirement for written informed consent was waived.

Study sample

We reviewed electronic pathology records from a single tertiary care center to identify consecutive adult patients (ages ≥ 18 years) who underwent liver transplantation as well as locoregional therapy between January 2011 and December 2015. Electronic medical records were searched to identify which patients underwent primary ablation for hepatocellular carcinoma prior to liver transplant. Inclusion criteria included multi-phase MRI no more than 90 days prior to treatment, at least one post-treatment multi-phase MRI performed no more than 90 days prior to transplantation, and no additional treatment to the lesion of interest prior to transplantation. The 90 day limit was based on the Organ Procurement and Transplantation Network policy of maximal allowable interval between imaging and liver transplantation [9].

Patients were excluded for the following reasons: 1) HCC treated at an outside facility with no pre-treatment MRI available; 2) no post-treatment MRI available prior to transplantation. Histopathology reports and MRIs for the included patients were reviewed and compared by a post-doctoral research fellow [MC] with two years of experience in image analysis with an abdominal radiology faculty member [LMBB] with 7 years of post-fellowship experience helping confirm difficult cases and questions.

Data Collection

Clinical and laboratory data were collected for each patient from the electronic medical record. Data included age, sex, presence of cirrhosis, underlying cause of chronic liver disease, serum total bilirubin, Model End-stage Liver Disease-Sodium score, time (days) between ablation and transplant, and Couinaud segment location of each embolized lesion.

Histopathologic data assessed at the time of liver explant by one of 7 surgical pathologists with 2, 7, 8, 12, 13, 22, and 30 years of experience were collected from the electronic pathology record. Collected data included tumor measurement, histopathologic diagnosis and tumor grade, and histopathologic tumor necrosis (≤99% necrotic, 100% necrotic).

MRI Technique

All imaging included in the study was consistent with LI-RADS v2018 and Organ Procurement and Transplantation Network technical requirements. Technical details regarding MRI systems and scan protocols are summarized in the Supplemental Materials section.

Treatment Technique

All lesions (n=53) were treated with either percutaneous, laparoscopic, or open surgical ablation. Technical details of the individual procedures are described in the Supplemental Materials section.

Image Analysis

Three faculty abdominal radiologists [BM, KAM, LMBB] with 2, 7, and 8 years of post-fellowship experience in abdominal MRI independently reviewed each patient's pre and

post-treatment MRI and evaluated them using the LI-RADS v2018 CT/MRI or Treatment Response Algorithm, respectively. Readers were blinded to all clinical data, and the reader who participated in initial review of MRIs for study inclusion [LMBB] underwent a 12 week washout period prior to image analysis. On pre-treatment MRI, each reader evaluated lesion size and the presence of major features including arterial phase hyperenhancement (APHE), washout, capsule, and threshold growth. Readers also reported pre-treatment LI-RADS category for each lesion based on LI-RADS major features only.

On post-treatment MRI, readers assessed lesion visibility, size of viable or equivocal disease, presence and absence of treatment response features, and treatment response category. Treatment response features included: expected post-treatment lesional hyperenhancement, categorized as LR-TR Nonviable; nodular, mass-like, or irregular thick tissue in or along the treated lesion with APHE, washout, and/or enhancement similar to pre-ablation, categorized as LR-TR Viable; or indeterminate enhancement, categorized as LR-TR Equivocal. If a lesion was no longer visible after treatment, it was categorized LR-TR Nonviable.

Statistical Analysis

Demographic, etiology, pre-treatment LI-RADS category, and pre-ablation lesion characteristics were summarized using mean with standard deviation (SD), median with range (min and max), or frequency with percentage (where appropriate). For patients with nodules diagnosed at different time points, summary statistics are based on data corresponding to each subject's earliest nodule diagnosis. The inter-reader agreement for pre and post-treatment MRI features and assessment categories were assessed using Fleiss' kappa with 95% confidence intervals obtained by bootstrapping. To determine the imaging features that are associated with histopathological tumor necrosis (≤99% necrosis, and 100%

necrosis), Pearson and Phi correlations (where appropriate), as well as Akaike Information Criterion values corresponding to exploratory simple logistic regression models were calculated using the "consensus" (2/3 reader consensus for yes/no responses and median value for numerical responses) value of the three readers for each of the imaging features of interest. For the exploratory simple logistic regression models, the independent observations assumption was relaxed. The consensus value for imaging features with a correlation coefficient of 0.30 or higher were used in respective logistic mixed effects models where subjects were treated as a random effect. The performance (sensitivity, specificity, positive predictive value, negative predictive value, and accuracy) of the TRA for each reader was approximated with respective generalized linear mixed effects models, where the reader was treated as a fixed effect, and the patient was treated as a random effect to account for multiple nodules on the same patient. Positive predictive value statistics were conducted under the condition where "Equivocal" assessments were treated as "Nonviable," and negative predictive value statistics were conducted under the condition where "Equivocal" assessments were treated as "Viable." The threshold for assessing statistical significance was set at alpha=0.05. Analyses were conducted using R: A language for statistical computing version 3.4.3 (R Foundation for Statistical Computing, Vienna, Austria) and SAS software version 9.4 (SAS Institute Inc, Cary, NC).

Results

Patient Characteristics

Of 72 adult patients who underwent liver transplantation for HCC and locoregional therapy, 42 received ablation therapy for hepatocellular carcinoma prior to liver transplant. 6 patients were excluded for the following reasons: HCC treated at an outside facility with no pre-treatment MRI available (n=2); no post-treatment MRI available prior to transplantation (n=4). In total, 36 adult patients (mean age 58±5 years, 32 men) with 53 lesions were included in the study, including 11% women (4/36, mean age 62 ± 5) and 89% men (32/36, mean age 57 ± 4.5).

All patients had cirrhosis with a majority (56%, 20/36) having both underlying Hepatitis C and prolonged alcohol consumption history. Median Model End-stage Liver Disease-Sodium score was 11 (range 6-26) and median BMI was 28.4 kg/m^2 (range 19-40.2 kg/m^2) prior to ablation. 19 nodules were treated with percutaneous radiofrequency ablation, 16 nodules were treated with intra-operative microwave laparoscopic ablation, 15 nodules were treated with percutaneous microwave ablation, and the final three nodules were treated with intra-operative radiofrequency ablation. Patient inclusion/exclusion is summarized in Figure 1, and pre-ablation patient characteristics are summarized in Table 1.

Consensus (2/3 readers) category assignments for pre-treatment MRI were: LI-RADS 4 (n=6) and LI-RADS 5 (n=47). 58% (31/53) of lesions had been ablated using microwave ablation, and the remaining 42% (22/53) had been ablated with radiofrequency ablation techniques. Median interval between locoregional treatment and transplant was 202 days (range 32-693). Of the patients with viable tumor on transplant histopathology, one demonstrated Grade 1 HCC, 14 demonstrated Grade 2 HCC, and seven demonstrated Grade 3 HCC. Of the six LR-4 lesions, three were confirmed as HCC on histopathology while three were completely necrotic and could not be confirmed.

Inter-reader Agreement

Inter-reader agreement for pre-treatment LI-RADS category (k=0.40 [95% CI: 0.15, 0.67]; p<0.01) was lower than the inter-reader agreement for post-treatment LR-TR category (k=0.71 [95% CI: 0.59, 0.84]; p<0.01). For post-treatment MRI features, readers were in highest agreement regarding the presence of nodular, mass-like, or irregular thick tissue in or along the treated lesions (NMLIT) with APHE (k=0.58 [95% CI: 0.40, 0.77]; p<0.01) followed by NMLIT with enhancement similar to pre-treatment (k=0.54; [95% CI: 0.28, 0.77]; p<0.01). Agreement was lowest for indeterminate hyperenhancement (k=0.26 [95% CI: -0.05, 0.60]; p<0.01). Inter-reader agreement is summarized in Supplemental Table 1.

Association Between Imaging Features and Histopathological Tumor Necrosis

The selection of median and consensus TRA features to assess respective associations with final histopathology are summarized in Supplemental Tables 2 and Table 3. 28% (15/53) of treated lesions were categorized as LR-TR Viable by consensus, 11 of the 15 demonstrating viable tumor on histopathology (Table 3). The four "false positive" tumors all demonstrated focal arterial hyperenhancement post-treatment. At histopathology, two demonstrated tumor along the ablation border, which was called viable by all three readers but were considered separate HCC lesions at pathological examination. One patient had an interval ablation of a separate HCC in an adjacent segment, which may have treated the viable disease prior to transplantation. The final patient had nodular peripheral enhancement which was considered viable tumor by two readers but was completely necrotic on histopathology. 11% (6/53) of treated lesions were categorized as LR-TR Equivocal by consensus; 5 of 6 were incompletely necrotic on histopathology. Finally, 60% (32/53) of the treated nodules were categorized as LR-TR Nonviable with 26 of the 53 completely necrotic

on histopathology. Figures 2 and 3 demonstrate examples of incompletely and completely ablated lesions, respectively.

The consensus assignment of post-treatment NMLIT with APHE, median viable or equivocally viable disease size, and the consensus assignment of NMLIT with enhancement similar to pre-treatment were selected for hypothesis testing (Supplemental Table 2). All three of the postulated mixed effects models yielded statistically significant results for the respective consensus imaging features of interest. Post-treatment NMLIT with APHE was the feature with the strongest consensus effect size (odds ratio: 142.86 [95% CI: 16.67,1250]; $p<0.01$) (Supplemental Table 3).

Diagnostic Performance for Predicting Histological Tumor Necrosis

Individual rater performance for predicting complete and incomplete tumor necrosis is summarized in Table 2 and Supplemental Table 4, accounting for clustering by subject. Sensitivity ranged from 40%-77% (15/22, 11/22, and 9/22) and specificity ranged from 85%-97% (26/31, 29/31, and 25/31) across readers when LR-TR Equivocal assessments were treated as "Nonviable". When LR-TR Equivocal assessments were treated as Viable, the sensitivity ranged from 81%-87% (16/22, 16/22, and 15/22) and the specificity ranged from 81%-85% (24/31, 25/31, and 24/31) across readers. Complete performance statistics are summarized in Supplemental Table 4.

Discussion

Since its release in 2011, LIRADS has attempted to standardize the assessment of HCC and facilitate therapeutic decision making. In our study sample composed of lesions treated with ablative techniques, the LI-RADS TRA performed well for predicting both complete (NPVs for residual tumor 89%-90%; 24/30, 25/31, and 24/31) and incomplete tumor necrosis (PPVs for residual tumor 70%-87%; 15/20, 11/13, and 9/15) when Equivocal assessments were treated as Viable or Nonviable, respectively (Supplemental Table 4). Of the imaging features, consensus NMLIT with APHE was the predictor of histopathological tumor necrosis with the largest effect size (odds ratio: 142.86 [95% CI: 16.67, 1250]; $p<0.01$). Although lesions rated as Nonviable or Viable accurately predicted their respective degree of tumor necrosis, 11% (6/53) were assessed as LR-TR Equivocal by consensus, and nearly all of these (5/6) were incompletely necrotic. This corresponds with known research; the assessment of tumor necrosis by imaging may not always be concordant with histopathologic necrosis, as incomplete necrosis is not uncommon in treated lesions that lack enhancement [(10)]. Treating Equivocal assessments as Viable (instead of Nonviable) assessments would not have substantially improved the PPV for incomplete necrosis (73%-76%; 16/23, 16/22, and 15/22), though it did reduce the range of performances between readers.

These findings contrast with recent work by Shropshire et al, who found slightly lower NPVs (81%-87%) but higher PPVs (86%-96%) for the TRA when applied to lesions that had undergone bland embolization (11). These differences could be related to the treatment modality, differences in years of reader experience, or performance of the MRIs in our study with a lower dose of gadolinium-based contrast agent (half the FDA-approved dose). Other studies have found strong performance of the TRA, though with little improvement over mRECIST (12).

Other studies have shown that APHE is the most important feature for imaging diagnosis in patients with HCC [7, 13, 14]. Recently, Burrel et al. validated APHE as a key marker of incomplete tumor necrosis after correlating imaging with histopathology [15]. The LI-RADS TRA similarly emphasizes APHE as well as including additional imaging features such as washout and enhancement similar to pre-treatment as markers of tumor viability. Our study shows that APHE is the dominant predictor of incomplete tumor necrosis, with other imaging features being complementary but less powerful predictors.

Inter-reader agreement is essential for the utilization of standardized reporting systems such as LI-RADS CT/MRI and the TRA. We observed greater inter-reader agreement for post-treatment LR-TR category (k=0.71 [95% CI: 0.59-0.84]; p<0.01) compared to pre-treatment LI-RADS category (k=0.40 [95% CI: 0.15, 0.67]; p<0.01). There was similar inter-reader agreement across most post-treatment MRI features (range k=0.54-0.58), except for "indeterminate hyperenhancement" and "washout," which demonstrated considerably lower reliability (k=0.26 and k=0.34 respectively). Indeterminate hyperenhancement was not considered a significant predictor of final histopathological necrosis, and clarification or elimination of this feature in future updates of the TRA may yield better results or reduce effort in applying the algorithm.

There were several limitations to our study, including a relatively small study sample of patients selected from a single tertiary care center. All included patients had MRIs for both pre- and post-treatment MRI. This reduced heterogeneity but limits the generalizability of our results to patients imaged with other modalities such as CT and contrast-enhanced ultrasound. The MRI protocol included use of half of the FDA-approved dose gadobenate dimeglumine intravenous injection during the study period, due to concerns at the time regarding nephrogenic systemic fibrosis. Several studies have reported half contrast dose as an

acceptable method for HCC imaging, but there is disagreement in the literature based on the potential for less avid lesional and hepatic parenchymal enhancement[16].

In conclusion, **our study shows that the LI-RADS v2018 Treatment Response algorithm performs well for lesions treated by ablation when observations are assessed as LR-TR Viable or Nonviable disease**. Nearly all lesions categorized as LR-TR Equivocal were incompletely necrotic on histopathology, which suggests that sensitivity for incompletely necrotic lesions might be improved if Equivocal lesions were, in fact, considered Viable. Arterial phase hyperenhancement was the most significantly associated predictor of histopathologic necrosis as well as the imaging feature with highest inter-reader agreement. Indeterminate hyperenhancement was not a significant predictor of tumor viability, and consideration should be given to clarifying or eliminating this feature. Further research is needed in a large multicenter multireader setting to more definitively assess the level of interreader reliability between radiologists with different training and experience levels. Also, our study only included lesions treated by ablation, and further work is needed to validate Treatment Response Algorithm performance treatment response using other locoregional therapies.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2018;68(6):394-424.

2. Ghouri YA, Mian I, Rowe JH. Review of hepatocellular carcinoma: Epidemiology, etiology, and carcinogenesis. Journal of carcinogenesis. 2017;16:1-.

3. Kielar A, Fowler KJ, Lewis S, Yaghmai V, Miller FH, Yarmohammadi H, et al. Locoregional therapies for hepatocellular carcinoma and the new LI-RADS treatment response algorithm. Abdominal Radiology. 2018;43(1):218-30.

4. Cescon M, Cucchetti A, Ravaioli M, Pinna AD. Hepatocellular carcinoma locoregional therapies for patients in the waiting list. Impact on transplantability and recurrence rate. Journal of Hepatology. 2013;58(3):609-18.

5. Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. Seminars in liver disease. 2010;30(1):52-60.

6. Patella F, Pesapane F, Fumarola EM, Emili I, Spairani R, Angileri SA, et al. CT-MRI LI-RADS v2017: A Comprehensive Guide for Beginners. Journal of clinical and translational hepatology. 2018;6(2):222-36.

7. Granata V, Fusco R, Avallone A, Catalano O, Filice F, Leongito M, et al. Major and ancillary magnetic resonance features of LI-RADS to assess HCC: an overview and update. Infectious Agents and Cancer. 2017;12(1):23.

8. Marrero JA, Kulik LM, Sirlin CB, Zhu AX, Finn RS, Abecassis MM, et al. Diagnosis, Staging, and Management of Hepatocellular Carcinoma: 2018 Practice Guidance by the American Association for the Study of Liver Diseases. Hepatology. 2018;68(2):723-50.

9. NETWORK OPAT. OPTN policies 2018 [Available from: https://optn.transplant.hrsa.gov/media/1200/optn_policies.pdf.

10. Chung W-S, Lee K-H, Park M-S, Lee YJ, Kwon J, Baek S-E, et al. Enhancement Patterns of Hepatocellular Carcinoma After Transarterial Chemoembolization Using Drug-Eluting Beads on Arterial Phase CT Images: A Pilot Retrospective Study. American Journal of Roentgenology. 2012;199(2):349-59.

11. Shropshire EL, Chaudhry M, Miller CM, Allen BC, Bozdogan E, Cardona DM, et al. LI-RADS Treatment Response Algorithm: Performance and Diagnostic Accuracy. Radiology. 2019;292(1):226-34.

12. Seo N, Kim MS, Park MS, Choi JY, Do RKG, Han K, et al. Evaluation of treatment response in hepatocellular carcinoma in the explanted liver with Liver Imaging Reporting and Data System version 2017. Eur Radiol. 2019.

13. Sharma P, Kalb B, Kitajima HD, Salman KN, Burrow B, Ray GL, et al. Optimization of single injection liver arterial phase gadolinium enhanced MRI using bolus track real-time imaging. Journal of Magnetic Resonance Imaging. 2011;33(1):110-8.

14. Shropshire EL, Chaudhry M, Miller CM, Allen BC, Bozdogan E, Cardona DM, et al. LI-RADS Treatment Response Algorithm: Performance and Diagnostic Accuracy. Radiology.0(0):182135.

15. Burrel M, Llovet JM, Ayuso C, Iglesias C, Sala M, Miquel R, et al. MRI angiography is superior to helical CT for detection of HCC prior to liver transplantation: An explant correlation. Hepatology. 2003;38(4):1034-42.

16. Homayoon B, Diwakar H, Strovski E, Bakshi D, Harris AC, Thoeni RF, et al. Half-dose gadobenate dimeglumine versus standard-dose gadodiamide in dynamic magnetic resonance imaging of non-cirrhotic livers: a retrospective intra-individual crossover comparison. Abdom Imaging. 2014;39(5):955-62.

Supplemental Materials

Imaging technique

MRI examinations were performed on either a 1.5T (Magnetom Avanto or Aera, Siemens Healthcare) or 3T (Magnetom Trio or Skyra, Siemens Healthcare) MRI system. 49 pretreatment scans were performed on a 1.5T, four on a 3T. 38 post-treatment scans were performed on a 1.5T, 15 on a 3T. A torso phased array coil was used in all studies.

All MRI examinations were performed to include a standard abdominal MRI protocol including: coronal T2-weighted half-Fourier single-shot fast spin-echo, coronal unenhanced T1-weighted nonfat-suppressed, axial unenhanced T1-weighted dual-echo in-phase/out-of-phase 2D gradient-echo (GRE), axial T2-weighted half-Fourier single-shot fast spin-echo sequences with fat suppression, and axial and coronal fat-suppressed 3D GRE sequences for dynamic contrast-enhanced study. All examinations were performed with half dose gadobenate dimeglumine intravenous injection (0.05 mmol/kg) using a power injection system (Spectris Solaris® EP, Medrad, Pittsburgh, PA) at a rate of 2 mL/s followed by a 20 mL saline bolus flush.

Treatment Technique

Ablation procedures were performed using one of three different approaches (percutaneous, laparoscopic or open surgical), and choice of ablation approach was based on performing clinician preference due to factors including tumor location and patient history. Percutaneous radiofrequency ablations were performed by the members of interventional radiology team under the direct supervision of a fellowship-trained attending interventional radiologist with 8, 10, and 32 years of post-fellowship experience. All patients underwent conscious sedation with midazolam (1-3 mg) and fentanyl (50-150 mcg). Tumor localization

and ablation were guided by trans-abdominal ultrasound, CT, or both. Ablation time varied between 5 – 15 min with temperatures reaching above 55 degrees Celsius. Ablation tracts were cauterized as the probes were removed.

All surgical ablations were performed by a team of hepatobiliary surgeons under the supervision of an abdominal transplant specialist with 20 years of post-fellowship experience. Patients received general endotracheal anesthesia and single dose of cefazolin (2 grams, IV) prior to the procedure. For laparoscopic approach, an infraumbilical incision was made to dissect and access the peritoneum, pneumoperitoneum was established, and 2-3 access ports were placed. Open ablation procedures were performed via a right subcostal incision. After visual or intra-operative ultrasound tumor localization, for microwave ablative technique, 2 – 3 Covidien microwave ablation probes (Medtronic, Dublin, Ireland) were used to perform 10 min 45-watt ablation under ultrasound guidance. Likewise, for radiofrequency ablative technique PR-20 NetWare ablation probes (Johnson & Johnson, New Brunswick, NJ) were used to perform 10 min 65-watt ablation under ultrasound guidance. When necessary to completely cover the lesion being treated, multiple overlapping ablations were performed, after which probes removed and tracts coagulated.

Tables

Table 1: Patient Characteristics Prior to Treatment

Patient Characteristics	**Values**
Age at time of ablation (years)	57 (47-67)
BMI at time of ablation	28.4 (19.0-40.2), 1 missing
Interval between ablation and transplantation (days)	202 (32-693)
Total Bilirubin (mg/dL)	1.5 (0.4-6.3), 12 missing
Model End-Stage Liver Disease-Sodium Score	11 (6-26), 16 missing
Underlying Etiology of Liver Disease	
Hepatitis C + Alcohol	55.6% (20/36)
Hepatitis C only	27.8% (10/36)
Alcohol only	5.6% (2/36)
Hepatitis B	2.8% (1/36)
Non-alcoholic steatohepatitis	8.3% (3/36)
Lesion LI-RADS Category at Baseline	
LI-RADS 4 (n=6)	11.3% (6/53)
LI-RADS 5 (n=47)	88.7% (47/53)
Pre-Embolization Lesion Characteristics by Consensus	
Median Size (mm)	21 (11-54)
Arterial Phase Hyperenhancement (APHE)	88.7% (47/53)
Washout	92.5% (49/53)
Capsule	81% (43/53)
Threshold Growth	34% (18/53)
Treatment Modality	
Microwave Ablation (Intra-operative)	30% (16/53)
Microwave Ablation (Interventional Radiology)	28% (15/53)
Radiofrequency Ablation (Intra-operative)	6% (3/53)
Radiofrequency Ablation (Interventional Radiology)	36% (19/53)

Note: Values are summarized as median (range) or proportion (counts) as appropriate.

Table 2: Reader performance for predicting incomplete necrosis based on the LR-TR Viable category (vs. Equivocal + Nonviable) and complete necrosis based on the LR-TR Nonviable category (vs. Equivocal + Viable), accounting for subject clustering, with 95% confidence intervals.

Assessment	Sensitivity	Specificity	Accuracy
LR-TR Viable (vs. Nonviable + Equivocal)			
Reader 1	15/22 (77%; 45%, 93%)	26/31 (88%; 63%, 97%)	41/53 82% (64%, 92%)
Reader 2	11/22 (53%; 22%, 82%)	29/31 (97%; 81%, 100%)	40/53 80% (61%, 91%)
Reader 3	9/22 (40%; 14%, 73%)	25/31 (85%; 57%, 96%)	34/53 66% (45%, 82%)
LR-TR Nonviable (vs. Viable + Equivocal)			
Reader 1	16/22 (87%; 50%, 98%)	24/31 (81%; 54%, 94%)	40/53 84% (63%, 94%)
Reader 2	16/22 (87%; 50%, 98%)	25/31 (85%; 59%, 96%)	41/53 86% (66%, 95%)
Reader 3	15/22 (81%; 40%, 96%)	24/31 (81%; 54%, 94%)	39/53 81% (59%, 93%)

Table 3: Post-treatment Lesion Characteristics by Consensus LI-RADS Category

Post-Treatment Lesion Characteristics by Consensus			
	LI-RADS Viable	LI-RADS Equivocal	LI-RADS Nonviable
Histopathologic Necrosis			
<99% necrotic	11	5	6
100% necrotic	4	1	26
Post-Treatment MRI Characteristics			
% with expected post-treatment lesional hyperenhancement	13% (2/15)	0%	0%
% NMLIT with APHE	87% (13/15)	33% (2/6)	0%
% NMLIT with Washout	27% (4/15)	0%	0%
% NMLIT with Enhancement similar to pre-ablation	40% (6/15)	0%	0%
% with indeterminate hyperenhancement	0%	50% (3/6)	0%
Mean Size			
Mean Whole Lesion Size (single longest measurement; mm)	23.3 ± 5.5	22.3 ± 5.4	22.2 ± 8.4
Mean Viable or Equivocally Viable Disease Size (mm)	17.4 ± 7.6	8.2 ± 5.4	0 ± 0

** NMLIT = Nodular, mass-like, or irregular thick tissue in or along the treated lesion

Supplemental Table 1: Inter-reader Agreement for Pre and Post-treatment MRI Features

Imaging Feature Inter-reader Agreement	**κ Value (95% CI)**	**P-value**
Pre-treatment		
LI-RADS category	0.40 (0.15, 0.67)	<0.01
Post-treatment		
NMLIT with arterial phase hyperenhancement **(APHE)**	0.58 (0.40, 0.77)	<0.01
NMLIT with **Washout**	0.34 (0.04, 0.66)	<0.01
NMLIT with **Enhancement similar to pre-treatment**	0.54 (0.28, 0.79)	<0.01
Indeterminate hyperenhancement	0.26 (-0.05, 0.60)	<0.01
LR-TR category	0.71 (0.59, 0.84)	<0.01
LR-TR category (where 'Equivocal' = 'Nonviable')	0.67 (0.51, 0.80)	<0.01
LR-TR category (where 'Equivocal' = 'Viable')	0.85 (0.73, 0.97)	<0.01

Supplemental Table 2: Selection of consensus imaging features to test for associations with histopathologic necrosis, using Akaike Information Criterion and p-values from univariable logistic regression models and correlations

Rank	Imaging Feature	Correlation Coefficient	Akaike Information Criterion	P-value
1	Nodular, mass-like, or irregular thick tissue in or along the treated lesion (NMLIT) WITH Arterial phase hyperenhancement (APHE)	-0.49	62.81	<0.01
2	Viable or Equivocally Viable Disease Size*	-0.45	64.36	<0.01
3	Nodular, mass-like, or irregular thick tissue in or along the treated lesion (NMLIT) WITH enhancement similar to pre-embolization	0.30	70.92	0.06
4	Perilesional geographic/perfusional hyperenhancement	-0.24	72.31	0.99
5	Expected post-treatment lesional hyperenhancement (thin rim)	0.22	73.26	0.12
6	Nodular, mass-like, or irregular thick tissue in or along the treated lesion (NMLIT) WITH washout	0.19	73.94	0.19
7	Indeterminate Enhancement	-0.13	75.12	0.38

*Phi correlations were calculated for all variables except for “disease size”, which used a Pearson correlation

Supplemental Table 3: Logistic mixed effects models testing the association between median size or consensus imaging features and histopathologic necrosis, treating subject as a random effect

Imaging Feature Model	Odds Ratio (95% CI)	P-value
Nodular, mass-like, or irregular thick tissue in or along the treated lesion (NMLIT) WITH Arterial phase hyperenhancement (APHE) (reference = Present)	142.86 (16.67, 1250)	<0.01
Median Viable or Equivocally Viable Disease Size	0.85 (0.77, 0.94)	<0.01
Nodular, mass-like, or irregular thick tissue in or along the treated lesion (NMLIT) WITH enhancement similar to pre-embolization (reference = Present)	333.33 (16.67, 5000)	<0.01

Supplemental Table 4: Reader performance for predicting incomplete necrosis based on the LR-TR Viable category (vs. Equivocal + Non-Viable) and complete necrosis based on the LR-TR Non-Viable category (vs. Equivocal + Viable), accounting for subject clustering, with 95% confidence intervals. Original proportions are also given.

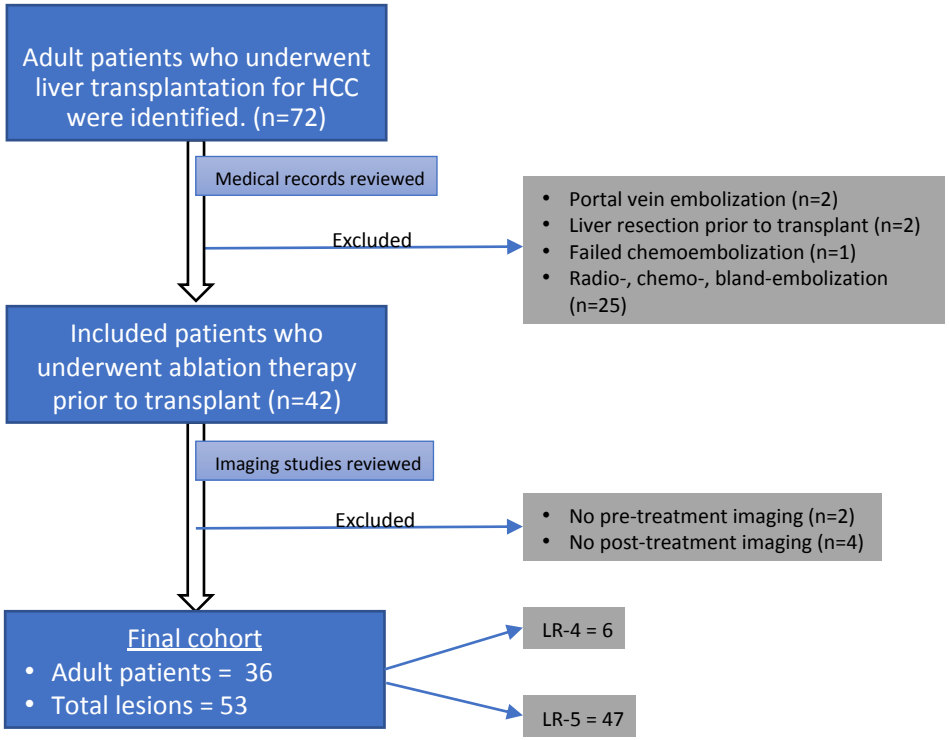
Assessment	Sensitivity	Specificity	PPV	NPV	Accuracy
LR-TR Viable (vs. Non-Viable + Equivocal)					
Reader 1	15/22 (77%; 45%, 93%)	26/31 (88%; 63%, 97%)	15/20 (80%; 39%, 96%)	26/33 (80%; 46%, 95%)	41/53 82% (64%, 92%)
Reader 2	11/22 (53%; 22%, 82%)	29/31 (97%; 81%, 100%)	11/13 (87%; 33%, 99%)	29/40 (77%; 45%, 93%)	40/53 80% (61%, 91%)
Reader 3	9/22 (40%; 14%, 73%)	25/31 (85%; 57%, 96%)	9/15 (70%; 24%, 95%)	25/38 (73%; 40%, 92%)	34/53 66% (45%, 82%)
LR-TR Non-Viable (vs. Viable + Equivocal					
Reader 1	16/22 (87%; 50%, 98%)	24/31 (81%; 54%, 94%)	16/23 73% (32%, 94%)	24/30 (90%; 52%, 99%)	40/53 84% (63%, 94%)
Reader 2	16/22 (87%; 50%, 98%)	25/31 (85%; 59%, 96%)	16/22 76% (33%, 95%)	25/31 (90%; 54%, 99%)	41/53 86% (66%, 95%)
Reader 3	15/22 (81%; 40%, 96%)	24/31 (81%; 54%, 94%)	15/22 74% (31%, 95%)	24/31 (89%; 52%, 98%)	39/53 81% (59%, 93%)

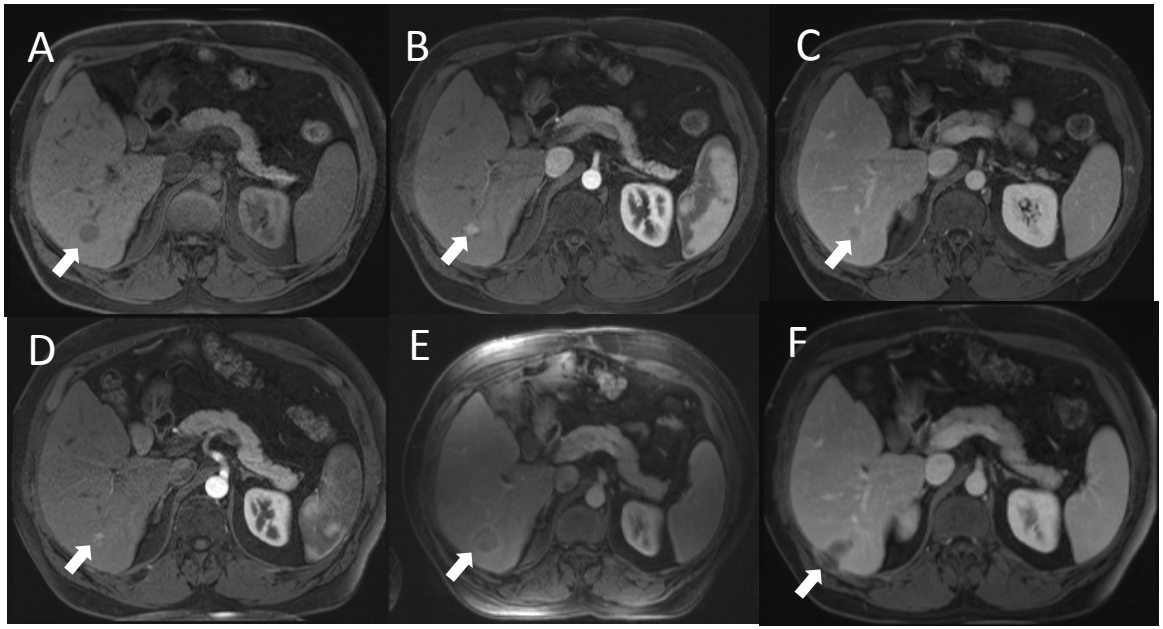
Figure Legends

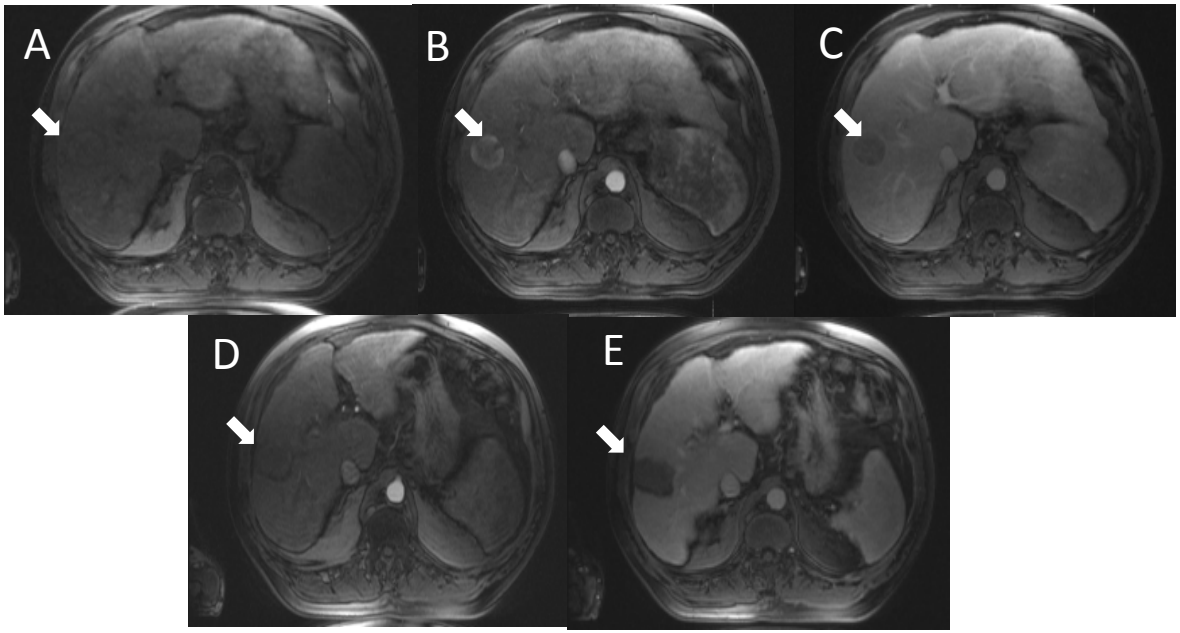
Figure 1. Study flow chart.

Figure 2. A-F. Images in a 55-year old man with cirrhosis due to Hepatitis C and alcohol consumption. Pre-treatment T1-weighted MRI demonstrates a 1.9 cm T1 hypointense lesion with arterial hyperenhancement in the right hepatic lobe (**A, B**), with washout in delayed venous phase **(C**), categorized LR-5, which was subsequently treated with microwave ablation. Post-treatment T1-weighted MRI in the hepatic arterial phase **(D)** demonstrates a 0.9 cm focus of persistent APHE, with perilesional rim enhancement in delayed venous phase (**E**) secondary to incomplete ablation (**F**), categorized LR-TR Viable. At histology this lesion was incompletely (≤99%) necrotic.

Figure 3. A-E. Images in a 49-year old man with cirrhosis due to Hepatitis C and alcohol consumption. Pre-treatment T1-weighted MRI demonstrates a 3.5 cm T1 hypointense lesion with arterial hyperenhancement in the right hepatic lobe (**A, B**), with washout in the delayed venous phase **(C),** categorized LR-5, which was subsequently treated with radiofrequency ablation**.** Post-treatment T1-weighted MRI in the hepatic arterial phase **(D)** shows no residual enhancement (arrow). The delayed venous phase **(E)** demonstrates post ablative changes with no washout (arrows), categorized LR-TR Nonviable. At histology this lesion was completely (100%) necrotic.


Adult patients who underwent liver transplantation for HCC were identified. (n=72)
Medical records reviewed
Excluded
Portal vein embolization (n=2)
Liver resection prior to transplant (n=2)
Failed chemoembolization (n=1)
Radio-, chemo-, bland-embolization (n=25)
Included patients who underwent ablation therapy prior to transplant (n=42)
Imaging studies reviewed
Excluded
No pre-treatment imaging (n=2)
No post-treatment imaging (n=4)
Final cohort
Adult patients = 36
Total lesions = 53
LR-4 = 6
LR-5 = 47


A
B
C
D
E
F


A
B
C
D
E

STARD 2015 Checklist

Section & Topic	No.	Item	✔ (N/A)
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	Y
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	Y
INTRODUCTION			
	3	Scientific & clinical background, including the intended use and clinical role of the index test	Y
	4	Study objectives and hypotheses	Y
METHODS			
Study design	5	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	Y
Participants	6	Eligibility criteria	Y
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	Y
	8	Where and when potentially eligible participants were identified (setting, location and dates)	Y
	9	Whether participants formed a consecutive, random or convenience series	Y
Test methods	10a	Index test, in sufficient detail to allow replication	Y
	10b	Reference standard, in sufficient detail to allow replication	Y
	11	Rationale for choosing the reference standard (if alternatives exist)	Y
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	Y
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	Y
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	Y
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	Y
Analysis	14	Methods for estimating or comparing measures of diagnostic accuracy	Y
	15	How indeterminate index test or reference standard results were handled	Y
	16	How missing data on the index test and reference standard were handled	Y
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	Y
	18	Intended sample size and how it was determined	Na
RESULTS			
Participants	19	Flow of participants, using a diagram	Y
	20	Baseline demographic and clinical characteristics of participants	Y
	21a	Distribution of severity of disease in those with the target condition	Y
	21b	Distribution of alternative diagnoses in those without the target condition	Na
	22	Time interval and any clinical interventions between index test and reference standard	Y
Test results	23	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	Y
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	Y
	25	Any adverse events from performing the index test or the reference standard	Na
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	Y
	27	Implications for practice, including the intended use and clinical role of the index test	Y
OTHER INFORMATION			
	28	Registration number and name of registry	Na
	29	Where the full study protocol can be accessed	Na
	30	Sources of funding and other support; role of funders	Na

*N/A stands for not applicable and may be a reasonable choice depending on the type of study performed