HHS Public Access

Author manuscript

Abdom Radiol (NY). Author manuscript; available in PMC 2020 June 01.

Published in final edited form as:

Abdom Radiol (NY). 2019 June ; 44(6): 2116–2132. doi:10.1007/s00261-019-01948-x.

Hepatocellular carcinoma (HCC) vs. non-HCC: accuracy and reliability of Liver Imaging Reporting and Data System v2018

Daniel R Ludwig, MD[1,*], **Tyler J Fraum, MD**[1], **Roberto Cannella, MD**[2], **David H Ballard, MD**[1], **Richard Tsai, MD**[1], **Muhammad Naeem, MD**[1], **Maverick LeBlanc, MD**[1], **Amber Salter, PhD**[1], **Allan Tsung, MD**[2], **Anup S Shetty, MD**[1], **Amir A Borhani, MD**[2], **Alessandro Furlan, MD**[2], and **Kathryn J Fowler, MD**[3]

[1]Mallinckrodt Institute of Radiology, Washington University School of Medicine, Saint Louis, MO, USA

[2]University of Pittsburgh Medical Center, Pittsburgh, PA, USA

[3]University of California San Diego, San Diego, CA, USA

Abstract

Purpose: The Liver Imaging Reporting and Data System (LI-RADS) was created to standardize the diagnostic criteria for hepatocellular carcinoma (HCC), and has undergone multiple revisions including a recent update in 2018 (v2018). The primary aim of this study was to determine the diagnostic performance and interrater reliability (IRR) of LI-RADS v2018 for distinguishing HCC from non-HCC primary hepatic malignancy in patients 'at-risk' for HCC. A secondary aim was to assess the impact of changes introduced in the v2018 diagnostic algorithm.

Methods: This retrospective study combined a 10-year experience of pathologically-proven primary liver malignancies from two large liver transplant centers. Two blinded readers independently evaluated each lesion and assigned a LI-RADS diagnostic category, additionally scoring all relevant imaging features. Changes in category based on the reader-provided features and the new v2018 criteria were assessed by a study coordinator.

Results: The final study cohort comprised 105 HCCs and 73 non-HCC primarily liver malignancies. LI-RADS had a high specificity for distinguishing HCC from non-HCC (89% and 90% for reader 1 and reader 2, respectively), and IRR was moderate to substantial for final LI-RADS category and most features. Revision of the LI-RADS v2018 diagnostic algorithm resulted in very few changes (5 [2.8%] and 3 [1.7%] for reader 1 and reader 2, respectively) in overall lesion classification.

[*]Corresponding Author Daniel R. Ludwig, MD, Mallinckrodt Institute of Radiology, Washington University School of Medicine, 510 S. Kingshighway Blvd, Campus Box 8131, Saint Louis, MO 63104, 314-362-5000, ludwigd@wustl.edu.

Conflict of interest: The authors have no potential conflicts of interest to disclose.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was waived by the institutional research committees at both institutions for this HIPAA compliant retrospective study.

Conclusion: LI-RADS diagnostic categories and features had moderate to substantial IRR and high specificity for distinguishing HCC from non-HCC primary liver malignancy. Revision of LI-RADS v2018 diagnostic algorithm resulted in reclassification of very few lesions.

Keywords

Interrater reliability; intrahepatic cholangiocarcinoma (iCCA); combined hepatocellular-cholangiocarcinoma (cHCC-CCA); cirrhosis; LI-RADS

Introduction

Hepatocellular carcinoma (HCC) is a leading cause of morbidity and mortality in patients with chronic liver disease, and it has become the most rapidly rising cause of cancer-related death in the United States [1]. Early detection of HCC allows effective management with locoregional therapy (LRT) or resection and may even permit cure by orthotopic liver transplantation (OLT) [2]. HCC is one of the few malignancies for which an imaging diagnosis is sufficient for directing management [3]. Accordingly, the Organ Procurement and Transplantation Network (OPTN) policy on organ allocation specifies that an imaging diagnosis and stage is sufficient to qualify for HCC model for end stage liver disease (MELD) exception points for liver transplantation. MELD exception points give patients priority on the liver transplantation waiting list. The primacy of imaging for this purpose diminishes the role of percutaneous biopsy, along with its attendant complications [4].

The American College of Radiology (ACR) developed the Liver Imaging Reporting and Data System (LI-RADS), which includes a diagnostic algorithm and standardized lexicon aimed at standardizing the imaging diagnosis of HCC [5]. LI-RADS has undergone multiple revisions, including recent updates in 2017 (v2017) and in 2018 (v2018) to achieve incorporation into the American Association for Study of Liver Diseases (AASLD) guidelines.

The 2017 update introduced new algorithms including ultrasound surveillance, contrast-enhanced ultrasound diagnosis, and treatment response assessment. Additionally, the main CT/MRI diagnostic algorithm was revised with specific criteria for LR-M (probable or definite malignancy, not specific for HCC). The LR-M category is intended to maximize sensitivity for diagnosing malignancy in general, while preserving specificity for diagnosing HCC (LR-5) [6, 7]. The rationale for creation of the LR-M category is that the risk factors for intrahepatic cholangiocarcinomas (iCCA) and combined hepatocellular-cholangiocarcinomas (cHCC-CCA) overlap with risk factors for HCC, such that all three subtypes of primary hepatic malignancy tend to afflict the same patient population. As patients with iCCA and cHCC-CCA have higher rates of recurrence and worse outcomes after OLT than patients with HCC [8, 9], it is important to diagnose HCC with high specificity and to diagnose iCCA and cHCC-CCA with high sensitivity.

To date, the evidence supporting the diagnostic criteria for LR-M primarily come from small single center cohorts reporting the imaging features of iCCA, atypical HCC, and cHCC-CCA, often in patients without prescribed risk factors for HCC [6, 10-13]. It is unknown whether iCCA and cHCC-CCA arising in the setting of cirrhosis exhibit these same imaging

features. Based on experience in prior studies, iCCA arising in cirrhosis have shown higher degrees of vascularity and potentially more closely resemble HCC [14]. Likewise, malignancies arising in a surveillance population (i.e., patients with cirrhosis) may be smaller at the time of first recognition. Hence, the ability of LI-RADS v2018 to differentiate HCC from other non-HCC primary liver malignancies in a high-risk cohort (i.e., the intended LI-RADS population) is unknown. The LI-RADS update in 2018 also affected the CT/MRI diagnostic algorithm, with changes to the definition of threshold growth and major feature requirements for categorization of LR-5 lesions measuring between 10 and 20 mm in size. In v2018, AASLD and LI-RADS are in alignment; however, the revisions create additional points of difference with OPTN [4]. The impact of these changes on lesion categorization has not yet been studied.

The primary aim of this study was to determine the diagnostic performance and interrater reliability (IRR) of LI-RADS for distinguishing HCC from non-HCC primary liver malignancies in patients stringently defined as ‘at-risk’ for HCC by LI-RADS criteria, using a cohort derived from a 10-year experience at two high volume transplantation centers. A secondary aim, given the recent update, was to assess the impact of the revisions to the diagnostic algorithm by determining the number of cases that changed LI-RADS categories with the modifications from v2017 to v2018 criteria.

Methods

Study Design

This Health Insurance Portability and Accountability Act (HIPAA)-compliant study was performed in a retrospective fashion at two large liver transplant centers. The institutional review board (IRB) at each center approved the protocol and waived the requirement for informed consent.

Study Cohort

Pathology served as the reference standard for diagnosis. The pathology databases from two large liver transplant centers were queried to identify all liver specimens logged between August 2007 and July 2017 with final diagnoses containing at least one of the following terms: hepatocellular carcinoma, cholangiocarcinoma, biphenotypic, and hepatocholangiocarcinoma. The terms biphenotypic and hepatocholangiocarcinoma were both included to identify all cHCC-CCA, as the pathology reports often used these terms interchangeably. cHCC-CCA was diagnosed based on morphologic features on routine histopathology with hematoxylin and eosin. Additional immunohistochemical testing was performed at the discretion of the interpreting pathologist, and supportive features included keratin 7 and 19 positivity, as well as expression of CD10 and pCEA within the biliary canaliculus [15, 16]. cHCC-CCA was considered a non-HCC primary liver malignancy in light of its association with worse post-OLT outcomes and the attendant controversy regarding the appropriateness of OLT for this indication [17]. The current CT and MRI protocols for both participating institutions have been previously published [6, 18]. Notably, given the 10-year interval from which eligible studies were identified, these protocols are

generally representative of our scanning techniques during this period, but do not account for minor year-to-year modifications.

Authors not involved in image interpretation reviewed the pathology reports to identify candidate liver masses. Lesions were excluded if the tissue received by pathology was deemed inadequate or if the final pathologic diagnosis was inconclusive. Specifically, cases of poorly differentiated carcinoma and adenocarcinoma (not otherwise specified) were excluded. If a patient had multiple lesions satisfying inclusion criteria, including those with more than one type of malignancy meeting inclusion criteria, the largest lesion was selected for LI-RADS evaluation, on the basis that the largest lesion usually drives clinical management. Furthermore, due to the high frequency of HCCs relative to other primary liver malignancies among at-risk patients, a relatively small subset of the identified HCCs was randomly selected (using a random number generator-based approach using cases occurring during the 2012-2015 interval) to limit the HCCs to one third of the total number of cases prior to risk factor assessment, though the proportion of HCCs rose with the subsequent exclusion of patients without LI-RADS risk factors (see below). The rationale was to facilitate a more robust analysis of non-HCC primary liver malignancies, with the understanding that the conclusions would be less generalizable given the resultant alterations to the relative frequencies of these lesion compared with those encountered in clinical practice. Note that a minority (less than 30%) of the identified patients were also included in a separate study unrelated to our current investigation [6]; images for all such patients were re-reviewed in a blinded fashion.

For the lesions satisfying the above criteria, clinical history and imaging was reviewed by authors uninvolved in image interpretation (DRL, TJF, RC, RT, MN, and ML) to identify patients eligible for LI-RADS assessment based on their risk factors. Per LI-RADS criteria, patients were required to have either cirrhosis or chronic hepatitis B viral infection. The presence of cirrhosis was assessed in a standardized fashion based on pathology, imaging, and laboratory results (Figure 1). Notably, patients with fibrosis but without cirrhosis were excluded [19]. Patients with chronic hepatitis B viral infection were included regardless of their cirrhosis status [20, 21]. Patients younger than 18 or with cirrhosis due to congenital hepatic fibrosis or other vascular disorders were excluded. Lesions were excluded if the patient did not undergo a liver-protocol dynamic contrast-enhanced (DCE) MRI or CT that satisfied the LI-RADS technical requirements [20, 21]. Additionally, lesions without a clear radiologic correlate or with multiple potential radiologic correlates were excluded. Any lesion that did not undergo imaging prior to LRT was also excluded. If multiple imaging studies were available, the study immediately before tissue acquisition (or immediately before the first LRT, if performed before tissue acquisition) was selected for LI-RADS assessment.

LI-RADS Assessment

Two fellowship-trained abdominal radiologists (KFJ and ASS) with 7 and 3 years of post-fellowship experience served as reader 1 (R1) and reader 2 (R2), respectively. Each reader had access to the patient’s age and gender but were otherwise blinded to clinical information, such as the original imaging interpretation and pathology results. Readers had

access to information from prior studies necessary to assess threshold growth. Information regarding whether a lesion was visualized as a discrete nodule on antecedent ultrasound was provided to the reviewer. Readers were directed to lesions by means of a series/image number, liver segment, and additional spatial identifying information if multiple lesions were present. Readers were asked to evaluate only the lesion of interest and not score additional lesions, as current versions of LI-RADS do not take multiplicity into account for assigning diagnostic categories. Each lesion was scored with respect to all major, LR-M, and ancillary features and assigned an overall LI-RADS diagnostic category. Readers applied tie-breaking rules and category adjustments according to LI-RADS methodology. Readers provided the final category based on v2017 criteria. A v2018 score was generated from the reader–provided data by the study coordinator using the new definition of threshold growth and new major feature criteria of LR-5 for 10-19 mm observations [20, 21]. Each reader subsequently reviewed the cases that changed category with v2018 and agreed with these changes. The LI-RADS v2018 score was used for all subsequent analyses, though the LI-RADS v2017 score was used to determine the number of lesions that changed scores between v2017 and v2018.

Statistical Analysis

Differences in mean between continuous variables were evaluated using the independent samples *t* test its non-parametric equivalent, Mann-Whitney U test, when applicable. Differences in frequencies of categorical variables, (e.g., LI-RADS features other than size) between HCC and non-HCC malignancy, were assessed using the Pearson χ^2 or Fisher exact test. The Cohen κ test was used to assess the IRR for categorical variables, whereas the intraclass correlation coefficient was used to assess the IRR for continuous variables. Agreement was scored as poor ($\kappa < 0.00$), slight ($\kappa = 0.00\text{-}0.20$), fair ($\kappa = 0.21\text{-}0.40$), moderate ($\kappa = 0.41\text{-}0.60$), substantial ($\kappa = 0.61\text{-}0.80$), or almost perfect ($\kappa = 0.81\text{-}1.00$) [22]. Confidence intervals for sensitivity and specificity were calculated according to the methods in Mercaldo *et al* [23]. Correction for multiple comparisons was performed using the methods of Benjamini and Hochberg to achieve a false discovery rate of 5% [24]; $p < 0.005$ was indicative of a significant difference. All statistical analyses were performed using R Studio (version 1.1.456, R Development Core Team, New Zealand).

Study Cohort

A total of 571 candidate liver specimens were obtained by query of the pathology database. Of these specimens, 393 were excluded based on our predefined criteria (Figure 2); the most common reasons for exclusion were ineligibility for LI-RADS assessment based on risk status (225 of 393, 57%), no liver-protocol CT or MRI (100 or 393, 25%), and no intraparenchymal mass on imaging (32 of 393, 8%). The final study cohort was comprised of 178 patients (Table 1). Average age was 61.9 ± 8.4 years, and patients were predominantly male (n = 138, 78%). Nearly all had cirrhosis (n = 174, 98%), with hepatitis C, non-alcoholic steatohepatitis (NASH), and alcohol-induced as the most common etiologies. A minority had chronic hepatitis B without cirrhosis (n = 4, 2%). In these 178 patients, there were 178 primary liver malignancies (Table 1). Despite enriching our population for non-HCC malignant lesions, a slight majority of the lesions were HCC (n = 105, 59%) as many non-HCC primary liver malignancies occurred in patients that did not

satisfy criteria for LI-RADS assessment (192 of 265, 72%). Pathologic diagnoses of HCC were based exclusively on resection (37 of 105, 35%) or explantation (68 of 105, 65%) pathology. In contrast, pathologic diagnoses of cHCC-CCA and iCCA were most commonly based on percutaneous biopsy (35 of 73, 48%). The interval between the imaging study for LI-RADS assessment to pathologic diagnosis was significantly higher in HCC versus non-HCC (189 ± 199 vs. 84 ± 132 days, $p < 0.001$), likely reflecting a higher rate of LRT in HCC versus non-HCC (64% vs. 21%, $p < 0.001$) and/or a lower rate of biopsy prior to treatment among HCCs. Imaging occurred with MRI and CT at similar rates in HCC and non-HCC (MRI: 85% vs. 82% for HCC and non-HCC, respectively; $p = 0.80$). Non-HCC malignancies were significantly larger than HCC (4.4 [1.5-14.0] vs 2.8 [1.4-19.0]; median [range]; $p < 0.001$).

Results

Diagnostic Performance of LI-RADS in the Prediction of HCC versus Non-HCC Primary Liver Malignancy

Figure 3 shows the number of lesions in each LI-RADS category stratified by reader (R1, R2) and pathologic diagnoses, while Table 2 shows the diagnostic performance of LI-RADS v2018 by reader for differentiating HCC from non-HCC malignancy. Specificity of LR-5 as a predictor of HCC was very high (89.0% and 90.4% for R1 and R2, respectively). Of the false positive LR-5 observations, the majority of these lesions were cHCC-CCAs (5 of 8, 63%, for R1; 6 of 7, 86%, for R2). Size of the false positive LR-5 observations was 3.1 ± 1.2 cm and 3.9 ± 3.3 cm for R1 and R2, respectively. When also considering LR-TIV (definitely due to HCC) as an imaging diagnosis of HCC, the specificity for HCC remained high (84.9% and 90.4% for R1 and R2, respectively). The sensitivities of LR-5 and LR-5 combined with LR-TIV (definitely due to HCC) as predictors of HCC were relatively limited (65.7% and 67.6%, respectively, for R1; 55.2% and 57.1%, respectively, for R2). Figure 4 shows a representative case of HCC scored by both readers as LR-5. Figure 5 shows an iCCA scored as LR-5 and LR-3 by R1 and R2, respectively, and Figure 6 shows a representative cHCC-CCA scored as LR-5 by both readers.

LR-M had moderate sensitivity for non-HCC malignancy (65.8% for R1, 72.6% for R2). However, combining LR-M with LR-TIV (may be due to non-HCC malignancy) for the prediction of non-HCC malignancy resulted in higher sensitivity (76.7% and 87.7% for R1 and R2, respectively). Interestingly, the combination of LR-M and LR-TIV (may be due to non-HCC malignancy) demonstrated a higher sensitivity for iCCA than for cHCC-CCA (88.0% vs 70.8% for R1, 92% vs 62.3% for R2). For the observations categorized as LR-M, 11 of 59 (18.6%) represented HCC for R1, and 26 of 79 (32.9%) represented HCC for R2. Figure 7 shows a representative case of cHCC-CCA scored by both readers as LR-M, whereas Figure 8 shows a representative case of HCC scored by both readers as LR-M. Very few observations were categorized as LR-3 (2 of 178, 1%, for both readers). Comparison of sensitivity and specificity of LI-RADS between MRI and CT was not performed, as there were so few observations scored on CT (31 of 178, 17%). Additionally, subgroup analysis by etiology of cirrhosis was not feasible due to the relatively limited numbers within each subgroup.

Interrater Reliability of LI-RADS Categories

Table 3 shows the IRR results for the LI-RADS categories. Agreement for overall LI-RADS category was moderate ($\kappa = 0.50$). Similarly, agreement on LR-5 or LR-TIV (definitely due to HCC) was moderate ($\kappa = 0.51$). However, agreement on LR-M or LR-TIV (may be due to non-HCC malignancy) versus other categories was substantial ($\kappa = 0.63$). Agreement on LR-5 versus LR-M or LR-TIV (i.e., observations that are probably or definitely malignant but not eligible for OPTN exception points) was also substantial (K = 0.64).

Change in LI-RADS Categorization with LI-RADS v2018

A total of 5 and 3 observations changed categories from LI-RADS v2017 to v2018 for R1 and R2, respectively (Figure 9). As expected, these category changes only impacted the LR-4 and LR-5 categories, and no LR-M or LR-TIV observations changed categories. For example, R1 scored a 5.2 cm observation with nonrim APHE and no additional major criteria that was new from a prior study four months earlier as LR-5 (per v2017); however, this appearance of a new lesion > 10 mm in diameter represents *sub*threshold growth according to v2018 and changed categories to LR-4 with v2018 (previously meeting criteria for threshold growth). In contrast, R2 felt that this lesion had been present on the prior study but had grown ≥ 50% in maximum diameter, thus corresponding to a score of LR-5 per v2017 and v2018. Nearly all the lesions that changed categories between v2017 and v2018 were HCCs, though one cHCC-CCA was re-categorized from LR-4 to LR-5 by R2. As such, the specificity of LR-5 as a predictor of HCC was minimally greater using v2017 as compared with v2018 for R2 (91.8% and 90.4%, respectively). Specificity of LR-5 as a predictor of HCC did not change for R1. As expected, there was no change in the diagnostic performance of LR-M between v2017 and v2018. A total of 17 and 4 lesions for R1 and R2, respectively, no longer met criteria for threshold growth using the revised criteria in v2018; however, a minority of these observations changed overall LI-RADS category as a result.

Frequency and Interrater Reliability of LI-RADS Major, LR-M, and Ancillary Features

Table 4 shows the frequencies of major features by reader among HCC versus non-HCC malignancies, with IRR analysis. Tumor in vein (TIV) demonstrated substantial agreement between readers, whereas the agreement was moderate for nonrim APHE, nonperipheral “washout”, enhancing “capsule”, and size (when assessed as 10-19 mm versus ≥ 20 mm). However, when size was analyzed in a continuous fashion, agreement was almost perfect (intraclass correlation coefficient 0.93; 95% confidence interval: 0.90, 0.95). Agreement for threshold growth was poor; however, only a small number of lesions met criteria for threshold growth (7 each for R1 and R2), and the confidence interval for this parameter was large. Interestingly, TIV was more common among non-HCC malignancies (significant difference for R1, $p = 0.001$; borderline significant difference for R2, $p = 0.005$).

Frequencies of LR-M features by reader among HCC versus non-HCC malignancies with IRR analysis are shown in Table 5. Targetoid mass, rim APHE, and delayed central enhancement were the most frequent LR-M criteria present, and agreement for these features was moderate to substantial. Peripheral “washout”, targetoid restriction, infiltrative appearance, marked diffusion restriction, and necrosis or severe ischemia had only slight to fair agreement, however these features were infrequently observed. Table 6 shows the

frequencies by reader with IRR analyses of ancillary features favoring malignancy but not specific for HCC, or favoring HCC in particular among HCC versus non-HCC malignancy. IRR analysis was not performed for ancillary features favoring benignity because very few malignant lesions in this study were scored as having such imaging features.

Discussion

The current study demonstrated strong diagnostic performance of LI-RADS v2018 for predicting HCC vs. non-HCC primary hepatic malignancy. Indeed, LR-5 and LR-5 combined with LR-TIV (definitely due to HCC) had very high specificities for predicting HCC, nearly 90% for both reviewers. These values are higher than those previously reported for LI-RADS v2014 [6], likely a result of revising the definition of the major feature APHE to exclude rim APHE in v2017, as a means of satisfying LR-5 criteria. As expected, most false positives were due to cHCC-CCA, likely attributable to the predominance of the hepatocellular component [25, 26]. Interestingly, a recent retrospective study suggested that cHCC-CCA scored as LR-5 rather than LR-M may indicate a better prognosis after curative surgery [27], supporting the hypothesis that the predominant phenotype drives the imaging appearance as well as the prognosis. Sensitivity of LR-5 for predicting HCC, on the other hand, was limited for both reviewers. This result is expected, as the LI-RADS algorithm was designed to maximize specificity for HCC at the expense of sensitivity for HCC, so as to avoid misallocation of transplant livers to patients falsely thought to have HCCs. The sensitivity was not significantly impacted with the revision in criteria for LR-5 with v2018, as will be discussed below.

Conversely, LR-M combined with LR-TIV (may be due to non-HCC malignancy) had a moderately high sensitivity for predicting non-HCC malignancy, 77% for R1 and 88% for R2. Interestingly, LR-M and LR-TIV (may be due to non-HCC malignancy) demonstrated a higher sensitivity for identifying iCCA than cHCC-CCA, likely owing to the overlap in histologic and imaging features between cHCC-CCA and HCC and again emphasizing the challenge that these lesions pose to imaging diagnostic accuracy [25, 26]. LR-M as a category should capture observations that are highly likely to represent malignancy, but do not have imaging features specific to HCC. Accordingly, categorization of an observation as LR-M conveys that a biopsy may be necessary to establish a definitive diagnosis, and thus sensitivity is desired over specificity in this setting. Indeed, up to 37% of LR-M lesions represent HCC in prior studies utilizing LI-RADS v2014 and v2017 [21] (slightly greater than the rate in our study; 19% for R1, 33% for R2). In other words, biopsying the occasional HCC that was incorrectly designated as LR-M is generally preferable to miscategorizing an iCCA or cHCC-CCA as LR-5, due to the higher rates of recurrence and worse outcomes in patients with iCCA and cHCC-CCA after OLT [8, 9].

The interrater reliability (IRR) of LI-RADS v2018 was moderate for the overall diagnostic category ($\kappa = 0.50$), similar to rates reported in prior studies on LI-RADS v2014 [6, 28]. It is possible that overall agreement is limited, at least in part, due to the subdivision of LR-TIV into three different subcategories, as was introduced in LI-RADS v2017. Further study is needed to evaluate the agreement specifically among these tumor in vein subcategories. Agreement on LR-5 versus LR-M or LR-TIV, however, was substantial, an important finding

as that latter group comprises observations that are probably or definitely malignant but not eligible for OPTN exception points. Agreement between readers for nearly all major features was moderate to substantial, again similar to values reported in prior studies [6, 28-31]. An interesting finding in our study was that agreement with respect to maximum lesion diameter, when assessed in a categorical fashion (i.e. < 10 mm, 10-19 mm, and ≥ 20 mm), was only moderate despite a very high IRR. Size is a major feature for defining HCC by imaging, and in practice these somewhat arbitrary size thresholds carry great significance for HCC staging and organ allocation. Agreement may be greater in practice for these categorical labels than in a research setting, where size measurements are not directly translated to transplantation eligibility. Agreement for threshold growth was poor; however, there were relatively few lesions that met criteria for threshold growth (seven for both R1 and R2), and the associated confidence intervals were large. Notably, one would expect agreement between readers to be high for threshold growth, given the strong IRR for size measurements noted in this and other studies [6, 28]. Agreement for LR-M features was moderate to substantial for the most frequently encountered features, including targetoid appearance, rim APHE, and delayed central enhancement. This finding is important, given that a solitary LR-M feature, when present, is sufficient for assignment of the LR-M category. The remaining LR-M features had only slight to fair agreement; however, these features were infrequently scored as present, and thus confidence intervals for these features tended to be large.

The secondary aim of the current study was to evaluate the potential impact of the recent LI-RADS v2018 changes on final diagnostic categorization. We found that there were very few changes in categories for our patient cohort, and thus there were few changes in the diagnostic accuracy of LI-RADS v2018 compared with v2017. The recent LI-RADS update serves to simplify the algorithm and to achieve concordance with the AASLD criteria for an imaging diagnosis of HCC. First, the definition of threshold growth was simplified to match that of OPTN and AASLD, requiring a ≥ 50% size increase of a mass in ≤ 6 months. Second, the criteria for LR-5 (definite HCC) were revised. LR-5g (LR-5-growth) and LR-5us (LR-5-ultrasound) categories were eliminated for simplicity. Furthermore, an observation 10-19 mm in size with nonrim APHE and non-peripheral “washout” now meets criteria for LR-5 (previously LR-4, or LR-5us if visible on antecedent ultrasound). An observation 10-19 mm in size with nonrim APHE and threshold growth is now denoted as LR-5 (previously LR-5g). Notably, there remains a discrepancy in categorization of 10-19 mm lesions with nonrim APHE and washout appearance; such observations meet AASLD/LI-RADS criteria for HCC but do not satisfy OPTN criteria for HCC exception points.

Patients were included in this study only if they met eligibility criteria for LI-RADS assessment, specifically the presence of cirrhosis or chronic hepatitis B. Determining eligibility for LI-RADS assessment remains a challenge in clinical practice, as many patients do not undergo biopsy for pathologic confirmation of cirrhosis. Imaging provides a relatively high degree of specificity for advanced liver disease but has limited sensitivity for detecting fibrosis and/or “early” cirrhosis [32, 33]. In patients with a liver lesion and equivocal morphologic findings of cirrhosis on imaging, biopsy of the lesion and/or background liver is often required to guide clinical management [34]. MR elastography may

prove useful in establishing a non-invasive diagnosis of cirrhosis and eligibility for LI-RADS assessment in selected patients [35]. Additionally, further study is warranted to determine whether LI-RADS assessment is valid in patients with fibrosis but without frank cirrhosis or in patients with risk factors for cirrhosis such as chronic hepatitis C or non-alcoholic fatty liver disease (NALFD).

Our study had several limitations, first and foremost its retrospective study design. However, the relative rarity of non-HCC primary hepatic malignancies among at-risk patients necessitated a retrospective design so as to include a sufficient number of cases. Another important limitation was the requirement for a pathology reference standard, which may have introduced selection bias, as our population was not reflective of all patients that are eligible for LI-RADS assessment. As such, the results of our analysis are certainly not as generalizable as the results from analysis of a prospective cohort. Additionally, our population of non-HCC primary hepatic malignancy included only iCCA and cHCC-CCA, and did not include other primary malignancies such as hepatic lymphoma or angiosarcoma, further limiting the generalizability of this study. Furthermore, our requirement for a definitive pathologic diagnosis (i.e. exclusion of poorly differentiated carcinoma and adenocarcinoma not otherwise specified) and adequate imaging evaluation may have led to an overestimation of the diagnostic accuracy. Moreover, the readers were not blinded to the study design, which may have introduced bias by influencing their likelihood of perceiving a major or LR-M feature, and assigning an overall score of LR-5 or LR-M rather than LR-3 or LR-4.

Patients with HCC in our study exclusively underwent resection or explant as a means of pathologic diagnosis, whereas the majority of patients with iCCA or cHCC-CCA underwent biopsy as their means of pathologic diagnosis. However, a diagnosis of HCC based on resection or explant reduces potential for sampling error introduced by biopsy, as a cholangiocellular component can certainly be missed when only a small component of the tissue from a lesion is analyzed. For this reason, it is possible that some patients with a diagnosis of iCCA actually had cHCC-CCA (if the hepatocellular component was missed due to sampling error); however, for the purpose of this study both diagnoses were considered non-HCC malignancy. In our study, most LR-TIV observations were non-HCC on pathology, potentially because patients with HCCs exhibiting typical imaging features in association with TIV may be treated for HCC without a pathologic diagnosis but may never undergo resection or explant due to the presence of the TIV on imaging (i.e., outside of Milan criteria). Additionally, non-HCC primary liver malignancies were over-represented in our population relative to their natural frequency, as our intent was to generate more robust data on the non-HCC malignancies to test the LR-M criteria in a rigorous fashion. This limitation precluded assessment of positive and negative predictive value as the ratio of HCC to non-HCC primary liver malignancies in our study did not reflect the relative frequencies encountered in clinical practice.

In conclusion, LI-RADS v2018 had a high specificity for distinguishing HCCs from non-HCC primarily liver malignancies, and IRR was moderate to substantial for overall LI-RADS category and most major, LR-M, and ancillary features. Revision and simplification

of LI-RADS with v2018 to achieve concordance with the AASLD resulted in very few changes in lesion classification.

Acknowledgments

Funding: No funding was obtained for this study. Dr. Ludwig receives salary support from the Training OPportunities in Translational Imaging Education and Research (TOP-TIER) Holden Thorp grant at Washington University. Dr. Ballard receives salary support from National Institutes of Health TOP-TIER grant T32-EB021955.

References

1. El-Serag HB (2011) Hepatocellular Carcinoma. N Engl J Med 365:1118–27. [PubMed: 21992124]
2. Mazzaferro V, Regalia E, Doci R, Andreola S, Pulvirenti A, Bozzetti F, et al. (1996) Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. N Engl J Med 334:693–700. [PubMed: 8594428]
3. Roberts LR, Sirlin CB, Zaiem F, Almasri J, Prokop LJ, Heimbach JK, et al. (2018) Imaging for the diagnosis of hepatocellular carcinoma: A systematic review and meta-analysis. Hepatology 67:401–21. [PubMed: 28859233]
4. Wald C, Russo MW, Heimbach JK, Hussain HK, Pomfret EA, Bruix J (2013) New OPTN/UNOS policy for liver transplant allocation: standardization of liver imaging, diagnosis, classification, and reporting of hepatocellular carcinoma. Radiology 266:376–82. [PubMed: 23362092]
5. Mitchell DG, Bruix J, Sherman M, Sirlin CB (2015) LI-RADS (Liver Imaging Reporting and Data System): Summary, discussion, and consensus of the LI-RADS Management Working Group and future directions. Hepatology 61:1056–65. [PubMed: 25041904]
6. Fraum TJ, Tsai R, Rohe E, Ludwig DR, Salter A, Nalbantoglu I, et al. (2017) Differentiation of hepatocellular carcinoma from other hepatic malignancies in patients at risk: diagnostic performance of the liver imaging reporting and data system version 2014. Radiology 286:158–72. [PubMed: 28853673]
7. Elsayes KM, Hooker JC, Agrons MM, Kielar AZ, Tang A, Fowler KJ, et al. (2017) 2017 version of LI-RADS for CT and MR imaging: an update. Radiographics 37:1994–2017. [PubMed: 29131761]
8. Vilchez V, Shah MB, Daily MF, Pena L, Tzeng CW, Davenport D, et al. (2016) Long-term outcome of patients undergoing liver transplantation for mixed hepatocellular carcinoma and cholangiocarcinoma: an analysis of the UNOS database. HPB (Oxford) 18:29–34. [PubMed: 26776848]
9. Sapisochin G, Fidelman N, Roberts JP, Yao FY (2011) Mixed hepatocellular cholangiocarcinoma and intrahepatic cholangiocarcinoma in patients undergoing transplantation for hepatocellular carcinoma. Liver Transpl 17:934–42. [PubMed: 21438129]
10. Joo I, Lee JM, Lee SM, Lee JS, Park JY, Han JK (2016) Diagnostic accuracy of liver imaging reporting and data system (LI-RADS) v2014 for intrahepatic mass-forming cholangiocarcinomas in patients with chronic liver disease on gadoxetic acid-enhanced MRI. J Magn Reson Imaging 44:1330–8. [PubMed: 27087012]
11. Asayama Y, Nishie A, Ishigami K, Ushijima Y, Takayama Y, Fujita N, et al. (2015) Distinguishing intrahepatic cholangiocarcinoma from poorly differentiated hepatocellular carcinoma using precontrast and gadoxetic acid-enhanced MRI. Diagn Interv Radiol. 21:96. [PubMed: 25698097]
12. Park HJ, Jang KM, Kang TW, Song KD, Kim SH, Kim YK, et al. (2016) Identification of imaging predictors discriminating different primary liver tumours in patients with chronic liver disease on gadoxetic acid-enhanced MRI: a classification tree analysis. Eur Radiol 26:3102–11. [PubMed: 26634931]
13. Fowler KJ, Potretzke TA, Hope TA, Costa EA, Wilson SR. LI-RADS M (LR-M): definite or probable malignancy, not specific for hepatocellular carcinoma (2018) Abdom Radiol (NY) 43:149–57. [PubMed: 28580538]
14. Xu J, Igarashi S, Sasaki M, Matsubara T, Yoneda N, Kozaka K, et al. (2012) Intrahepatic cholangiocarcinomas in cirrhosis are hypervascular in comparison with those in normal livers. Liver Int 32:1156–64. [PubMed: 22417172]

15. Bosman FT, Carneiro F, Hruban RH, Theise ND (2010) WHO classification of tumours of the digestive system. Lyon, France: IARC Press.
16. Brunt E, Aishima S, Clavien PA, Fowler K, Goodman Z, Gores G, et al. (2018) cHCC-CCA: Consensus terminology for primary liver carcinomas with both hepatocytic and cholangiocytic differentation. Hepatology 68:113–126. [PubMed: 29360137]
17. Magistri P, Tarantino G, Serra V, Guidetti C, Ballarin R, Di Benedetto F (2017) Liver transplantation and combined hepatocellular-cholangiocarcinoma: Feasibility and outcomes. Dig Liver Dis 49:467–70. [PubMed: 28258929]
18. Furlan A, Almusa O, Yu RK, Sagreiya H, Borhani AA, Bae KT, Marsh JW (2018) A radiogenomic analysis of hepatocellular carcinoma: association between fractional allelic imbalance rate index and the liver imaging reporting and data system (LI-RADS) categories and features. Br J Radiol 91:1086.
19. Suk KT, Kim DJ (2015) Staging of liver fibrosis or cirrhosis: The role of hepatic venous pressure gradient measurement. World J Hepatol 7:607. [PubMed: 25848485]
20. American College of Radiology. CT/MRI LI-RADS v2017 Core. https://www.acr.org/-/media/ACR/Files/RADS/LI-RADS/LIRADS_2017_Core.pdf (accessed 25 Jul 2018).
21. American College of Radiology. CT/MRI LI-RADS v2018 Core. https://www.acr.org/-/media/ACR/Files/RADS/LI-RADS/LI-RADS-2018-Core.pdf (accessed 25 Jul 2018).
22. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–74. [PubMed: 843571]
23. Mercaldo ND, Lau KF, Zhou XH (2007) Confidence intervals for predictive values with an emphasis to case–control studies. Stat Med 26:2170–83. [PubMed: 16927452]
24. Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Statist 29:1165–88.
25. Bergquist JR, Groeschl RT, Ivanics T, Shubert CR, Habermann EB, Kendrick ML, et al. (2016) Mixed hepatocellular and cholangiocarcinoma: a rare tumor with a mix of parent phenotypic characteristics. HPB (Oxford) 18:886–92. [PubMed: 27546172]
26. Fowler KJ, Sheybani A, Parker RA III, Doherty S, M Brunt E, Chapman WC, et al. (2013) Combined hepatocellular and cholangiocarcinoma (biphenotypic) tumors: imaging features and diagnostic accuracy of contrast-enhanced CT and MRI. AJR Am J Roentgenol 201:332–9. [PubMed: 23883213]
27. Jeon SK, Joo I, Lee DH, Lee SM, Kang HJ, Lee KB, et al. (2018) Combined hepatocellular cholangiocarcinoma: LI-RADS v2017 categorisation for differential diagnosis and prognostication on gadoxetic acid-enhanced MR imaging. Eur Radiology. doi: 10.1007/s00330-018-5605-x. [Epub ahead of print]
28. Fowler KJ, Tang A, Santillan C, Bhargavan-Chatfield M, Heiken J, Jha RC, et al. (2018) Interreader Reliability of LI-RADS version 2014 algorithm and imaging features for diagnosis of hepatocellular carcinoma: a large international multireader study. Radiology 286:173–85. [PubMed: 29091751]
29. Ehman EC, Behr SC, Umetsu SE, Fidelman N, Yeh BM, Ferrell LD, et al. (2016) Rate of observation and inter-observer agreement for LI-RADS major features at CT and MRI in 184 pathology proven hepatocellular carcinomas. Abdom Radiol (NY) 41:963–9. [PubMed: 27193793]
30. Davenport MS, Khalatbari S, Liu PS, Maturen KE, Kaza RK, Wasnik AP, et al. (2014) Repeatability of diagnostic features and scoring systems for hepatocellular carcinoma by using MR imaging. Radiology 272:132–42. [PubMed: 24555636]
31. Becker AS, Barth BK, Marquez PH, Donati OF, Ulbrich EJ, Karlo C, et al. (2017) Increased interreader agreement in diagnosis of hepatocellular carcinoma using an adapted LI-RADS algorithm. Eur J Radiol 86:33–40. [PubMed: 28027763]
32. Horowitz JM, Venkatesh SK, Ehman RL, Jhaveri K, Kamath P, Ohliger MA, et al. (2017) Evaluation of hepatic fibrosis: a review from the society of abdominal radiology disease focus panel. Abdom Radiol (NY) 42:2037–53. [PubMed: 28624924]
33. Kudo M, Zheng RQ, Kim SR, Okabe Y, Osaki Y, Iijima H, et al. (2008) Diagnostic accuracy of imaging for liver cirrhosis compared to histologically proven liver cirrhosis. Intervirology 51:17–26. [PubMed: 18544944]

34. Heimbach JK, Kulik LM, Finn RS, Sirlin CB, Abecassis MM, Roberts LR, et al. (2018) AASLD guidelines for the treatment of hepatocellular carcinoma. Hepatology 67:358–80. [PubMed: 28130846]

35. Venkatesh SK, Yin M, Ehman RL (2013) Magnetic resonance elastography of liver: technique, analysis, and clinical applications. J Magn Reson Imaging 37:544–55. [PubMed: 23423795]

36. Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, et al. (2006) Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. Hepatology 43:1317–25. [PubMed: 16729309]

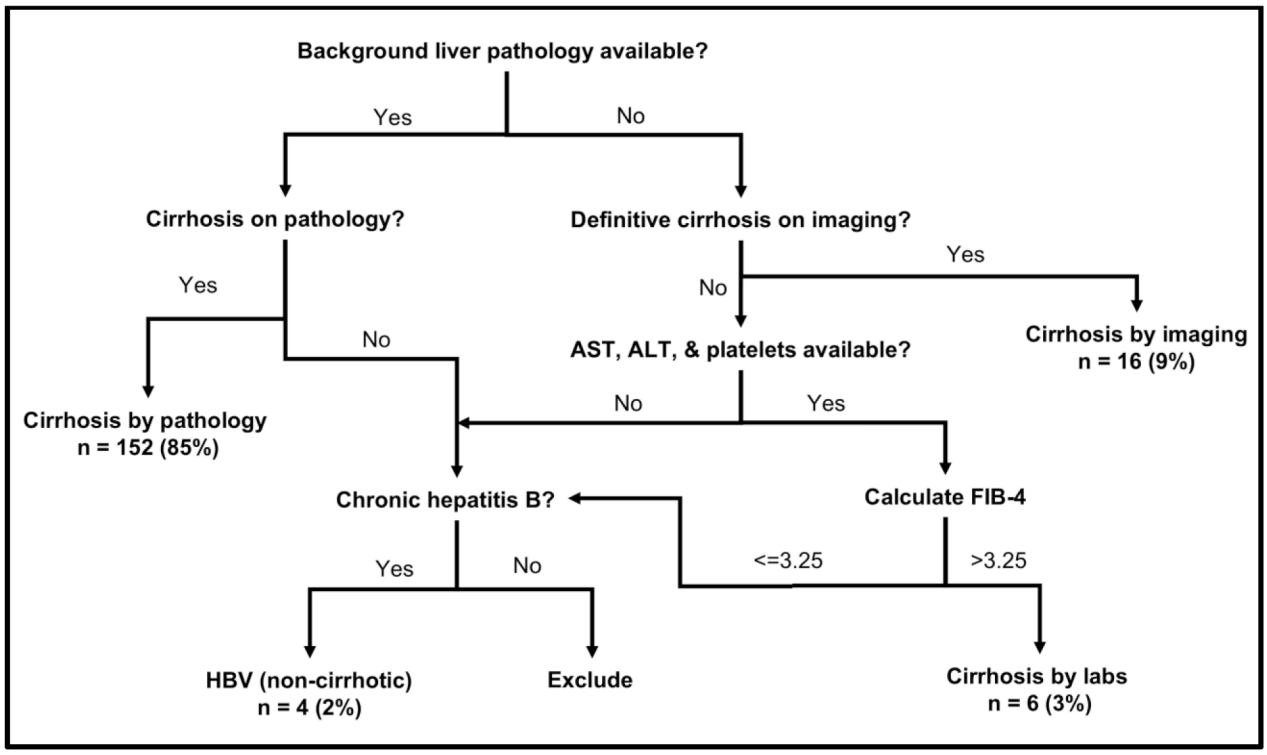

Fig 1.

Algorithm for LI-RADS Eligibility Assessment. Patients were eligible if they had cirrhosis by pathology, imaging, or laboratory analysis. If background liver tissue was available to the interpreting pathologist, this assessment was preferentially used for risk status assessment. Notably, patients with fibrosis but without cirrhosis were excluded [20, 21]. If no background liver tissue was available (i.e., pathology specimen included mass only), patients were considered cirrhotic if the interpreting radiologist felt that the liver exhibited unequivocal surface nodularity. Because the assessment of cirrhosis by imaging provides a high degree of specificity (77.4-99%) but somewhat limited sensitivity (59-93%) [32], If the patient did not have cirrhosis on imaging, laboratory values (if available) were used to calculate a FIB-4 score. Patients with a FIB-4 greater than 3.25 were considered to have cirrhosis, as a FIB-4 > 3.25 specificity of 97% for the detection of advanced fibrosis [36]. Patients with chronic hepatitis B viral infection were included regardless of their cirrhosis status [20, 21]. Patients younger than 18 or with cirrhosis due to congenital hepatic fibrosis or other vascular disorders were excluded.

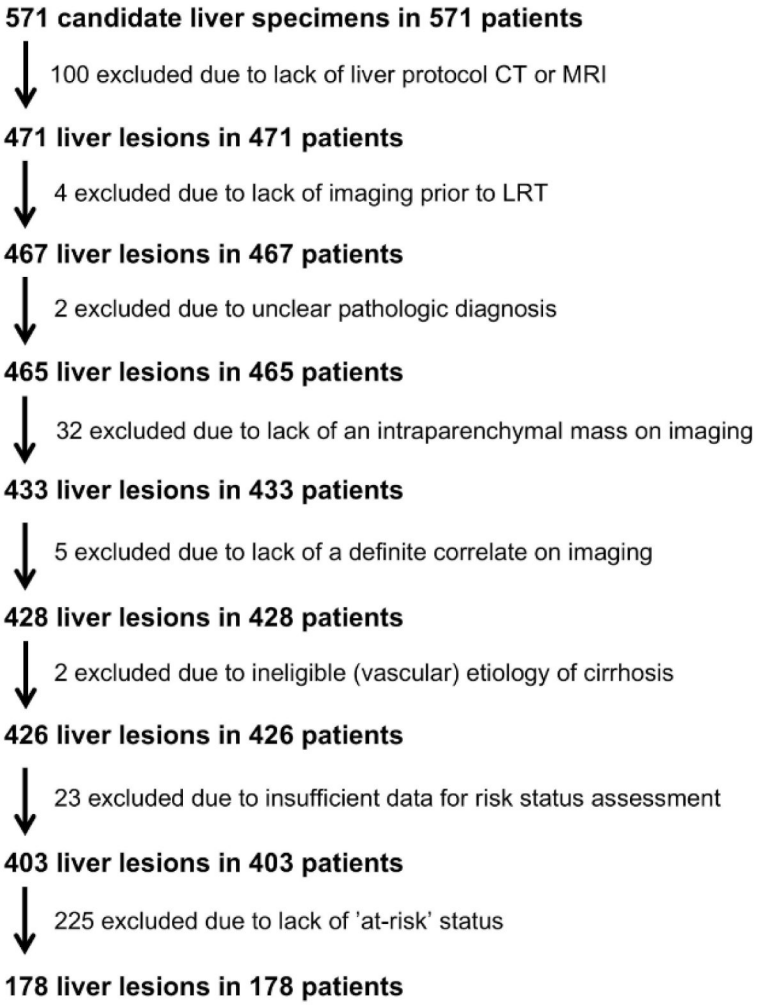

Fig 2.

Flowchart demonstrating number of lesions excluded for each of our predefined criteria. Abbreviations: CT – computed tomography; LRT – locoregional therapy; MRI – magnetic resonance imaging.

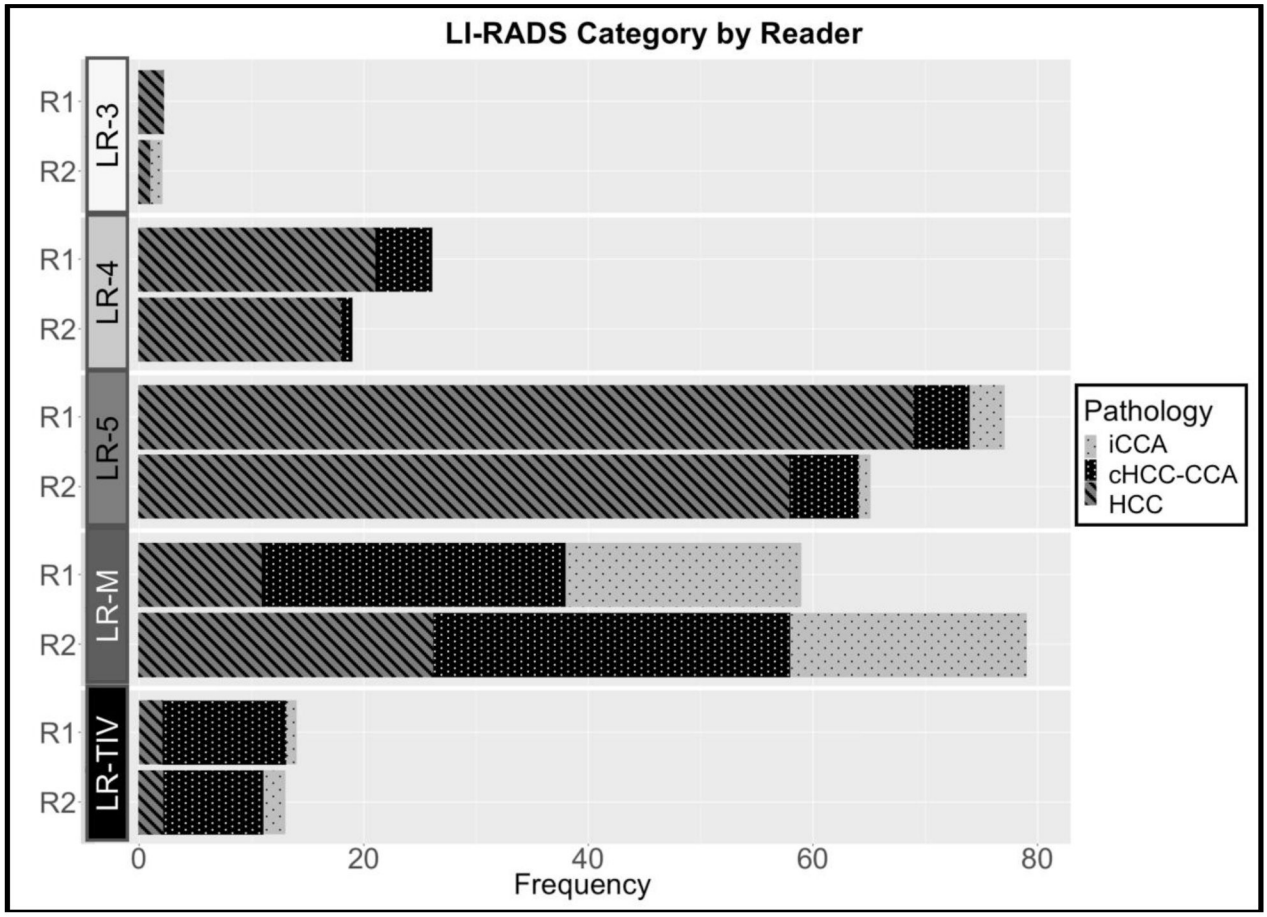

Fig 3.

Number of lesions in each LI-RADS category stratified by reader and pathologic diagnosis. No lesions were classified as LR-1 or LR-2. LR-TIV was further scored as definitely due to HCC, probably due to HCC, or may be due to non-HCC malignancy according to LI-RADS v2018 (not shown).

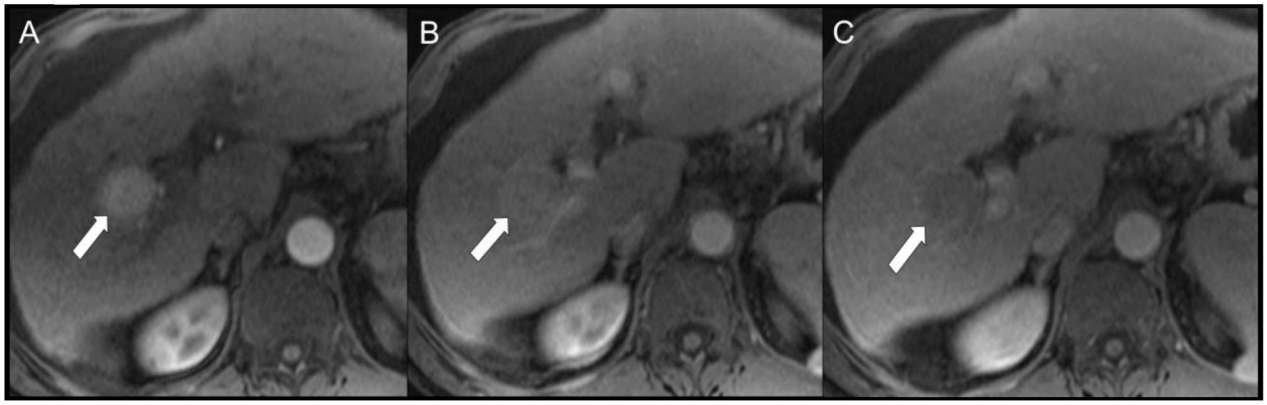

Fig 4.

Representative HCC scored as LR-5 by both readers. Patient is a 57-year-old man with cirrhosis secondary to hepatitis C virus undergoing MR screening examination. Arterial (A), portal venous (B), and delayed (C) post-contrast sequences obtained after gadoversetamide (OptiMARK) administration demonstrate an observation with nonrim APHE (A, arrow) at the junction of segments of 5 and 6 that demonstrated nonperipheral “washout” (B and C, arrows) and an enhancing “capsule” (C, arrow). Both readers scored all these features present, and categorized this observation at LR-5 using both v2017 and v2018.

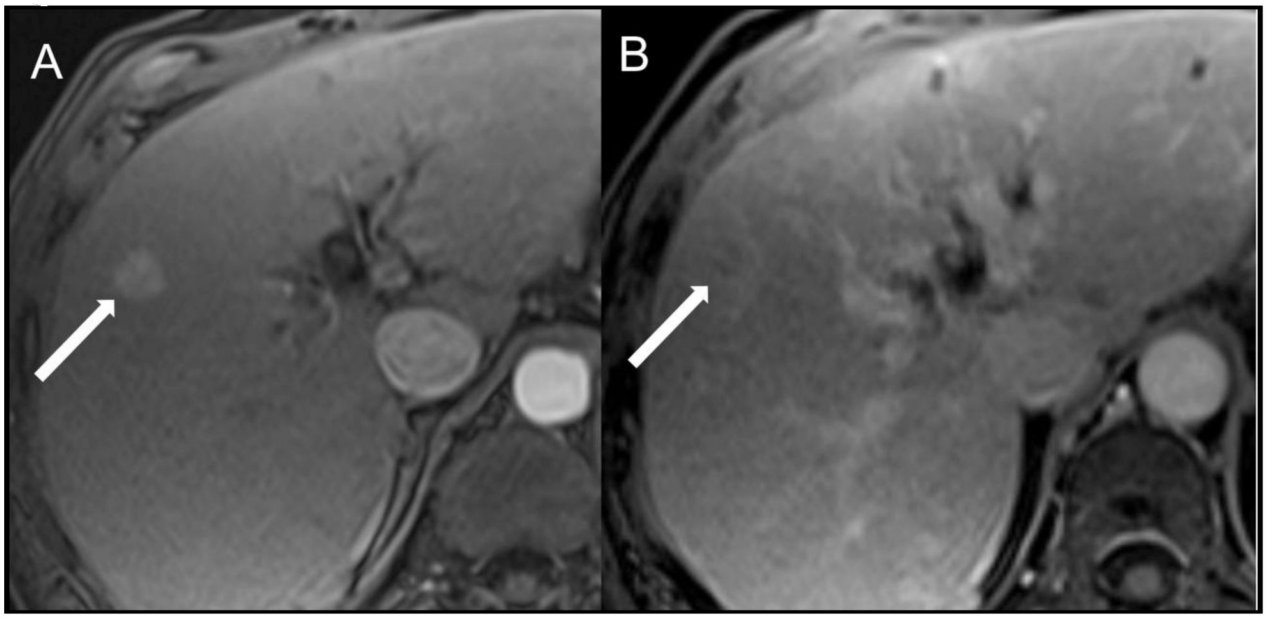

Fig 5.

Example of an intrahepatic cholangiocarcinoma (iCCA) with a hypervascular appearance. Patient is a 56-year-old man with cirrhosis secondary to hepatitis C who underwent screening MR examination. Arterial (A) and delayed phase (B) images after gadoversetamide (OptiMARK) administration demonstrate an observation measuring approximately 15 mm at the junction of segments 5 and 8 with non-rim APHE (A, arrow) and an ill-defined rounded area of decreased intensity with peripheral hyperintensity on the delayed phase (B, arrow). R1 scored this lesion as having nonperipheral “washout” and an enhancing “capsule,” whereas R2 did not score these features as present. Overall LI-RADS categories were LR-5 and LR-3 for R1 and R2, respectively.

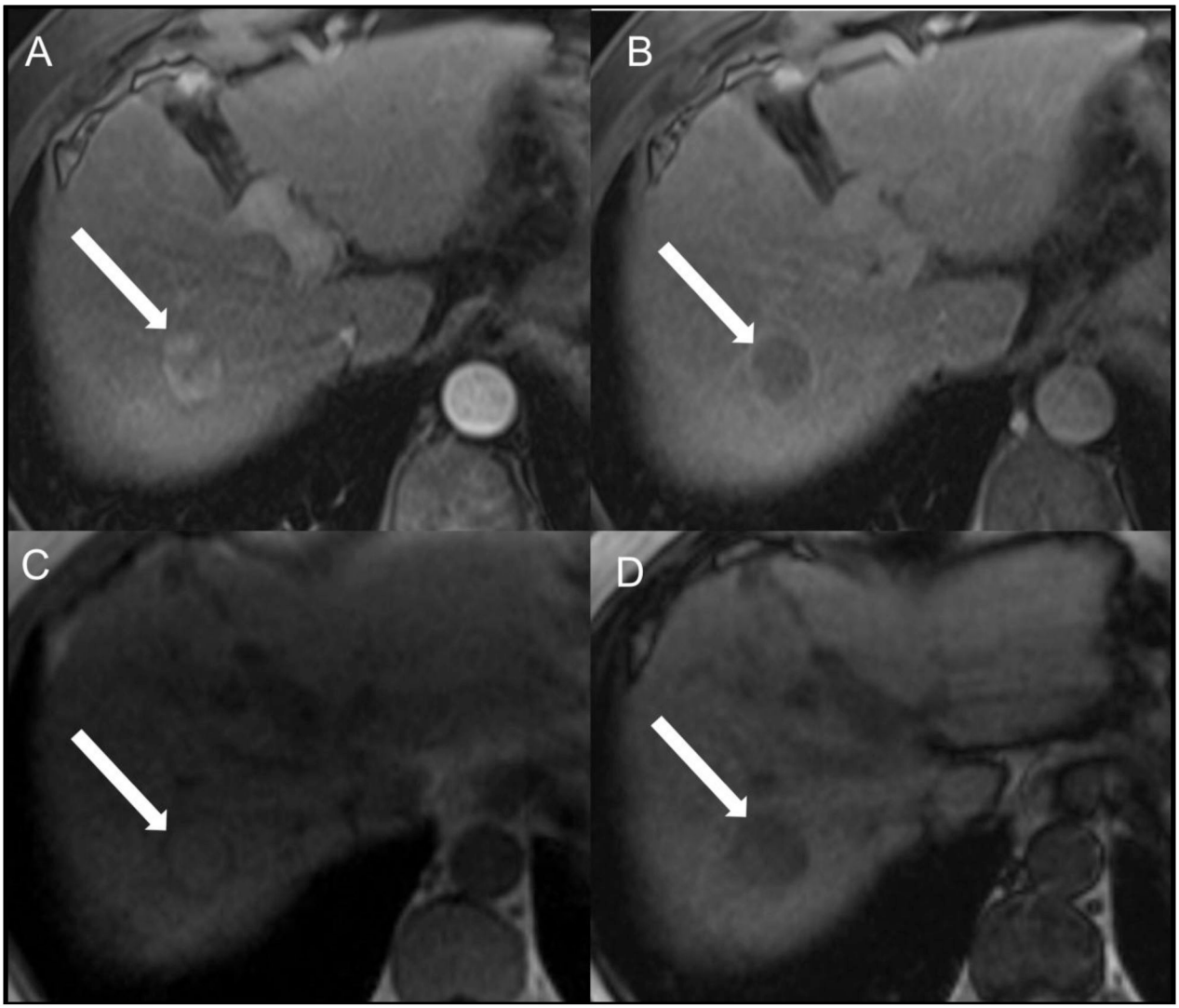

Fig 6.

Representative cHCC-CCA scored as LR-5 by both readers. Patient is a 65-year-old man with cirrhosis secondary to hepatitis B virus who underwent evaluation after a identification of a solid nodule on screening ultrasound. Contrast-enhanced MR in the arterial (A) and delayed-phase (B) after gadoversetamide (OptiMARK) administration demonstrate an observation at the junction of segments 7 and 8 (arrows) with nonrim APHE, nonperipheral “washout”, and an enhancing “capsule.” Additionally, in-phase (C) and out-of-phase (D) images demonstrate signal loss on out-of-phase acquisition, consistent with microscopic fat within the lesion, greater than adjacent liver. Overall LI-RADS category was LR-5 for both readers.

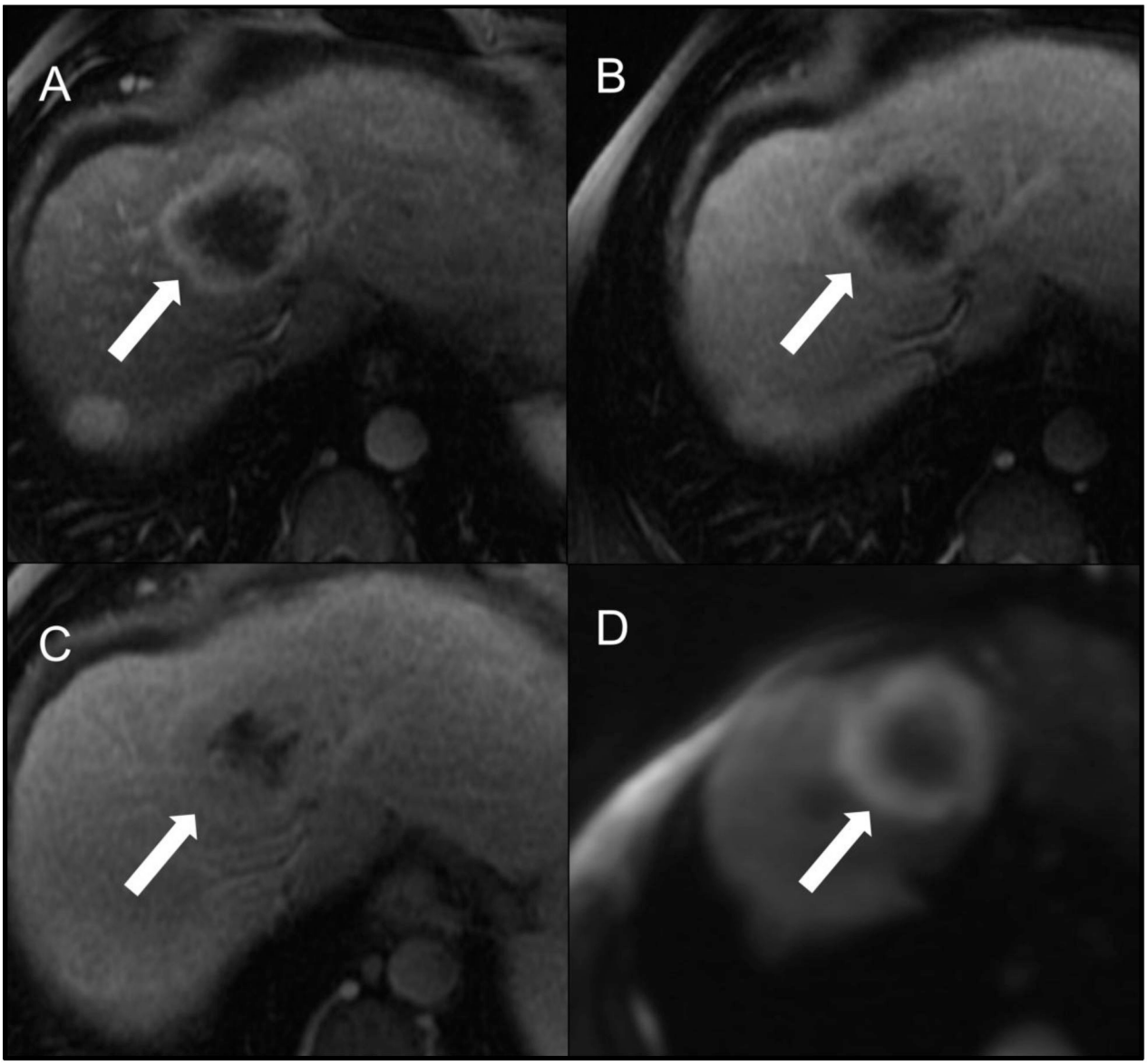

Fig 7.

Representative cHCC-CCA scored as LR-M by both readers. Patient is a 63-year-old male with alcoholic cirrhosis undergoing MR examination following an ultrasound screening examination revealing a segment 8 intrahepatic lesion (not shown). Arterial (A), portal venous (B), delayed (C) post-contrast sequences and diffuse-weighted imaging (D) after gadobenate dimeglumine (MultiHance) administration demonstrate a segment 8 observation measuring approximately 6 cm in size (arrows). Both study readers agreed on the presence of targetoid appearance (including targetoid restriction) and rim APHE. R1 categorized the lesion as having peripheral “washout” and a necrotic appearance whereas R2 did not; R2 categorized the lesion as having delayed central enhancement whereas R1 did not. Both readers categorized this observation as LR-M using LI-RADS v2017 and v2018.

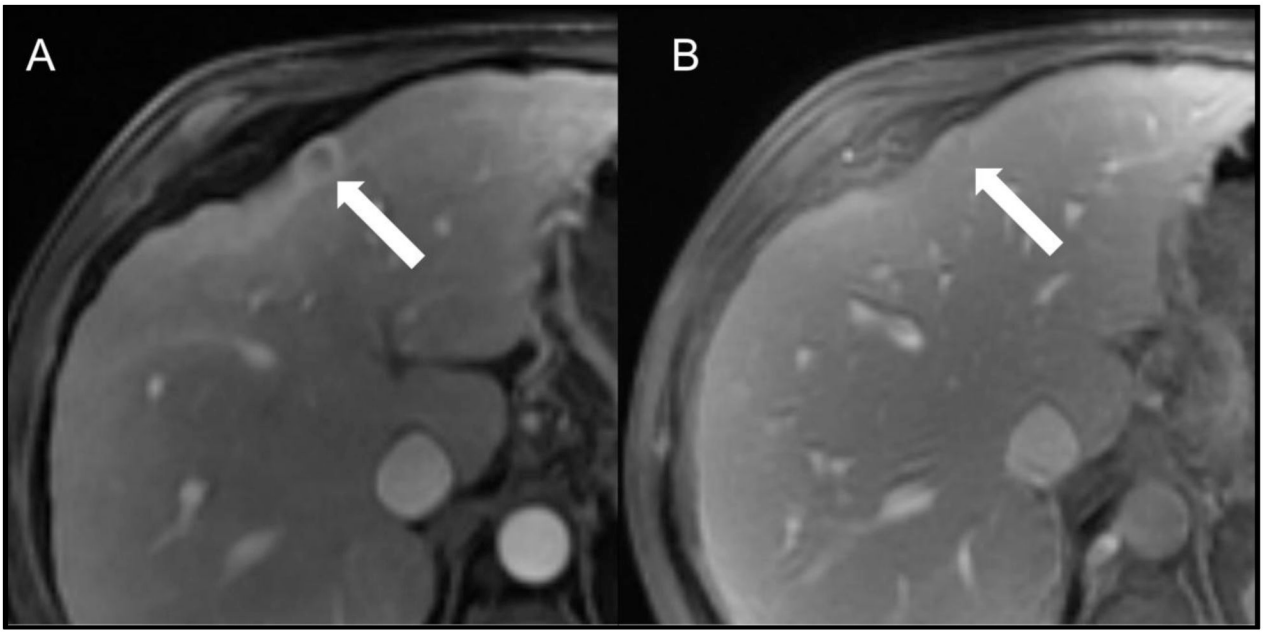

Fig 8.

Example HCC scored as LR-M by both readers. Patient is a 60-year-old man with cirrhosis secondary to hepatitis C virus who underwent screening MR examination. Contrast-enhanced MR in the arterial (A) and delayed-phase (B) images after gadoxetate disodium (Eovist) demonstrate an observation in segment 4A (arrows) with a targetoid appearance and rim APHE. R2 additionally scored the observation as having peripheral “washout” and delayed central enhancement. Overall LI-RADS category was LR-M for both readers.

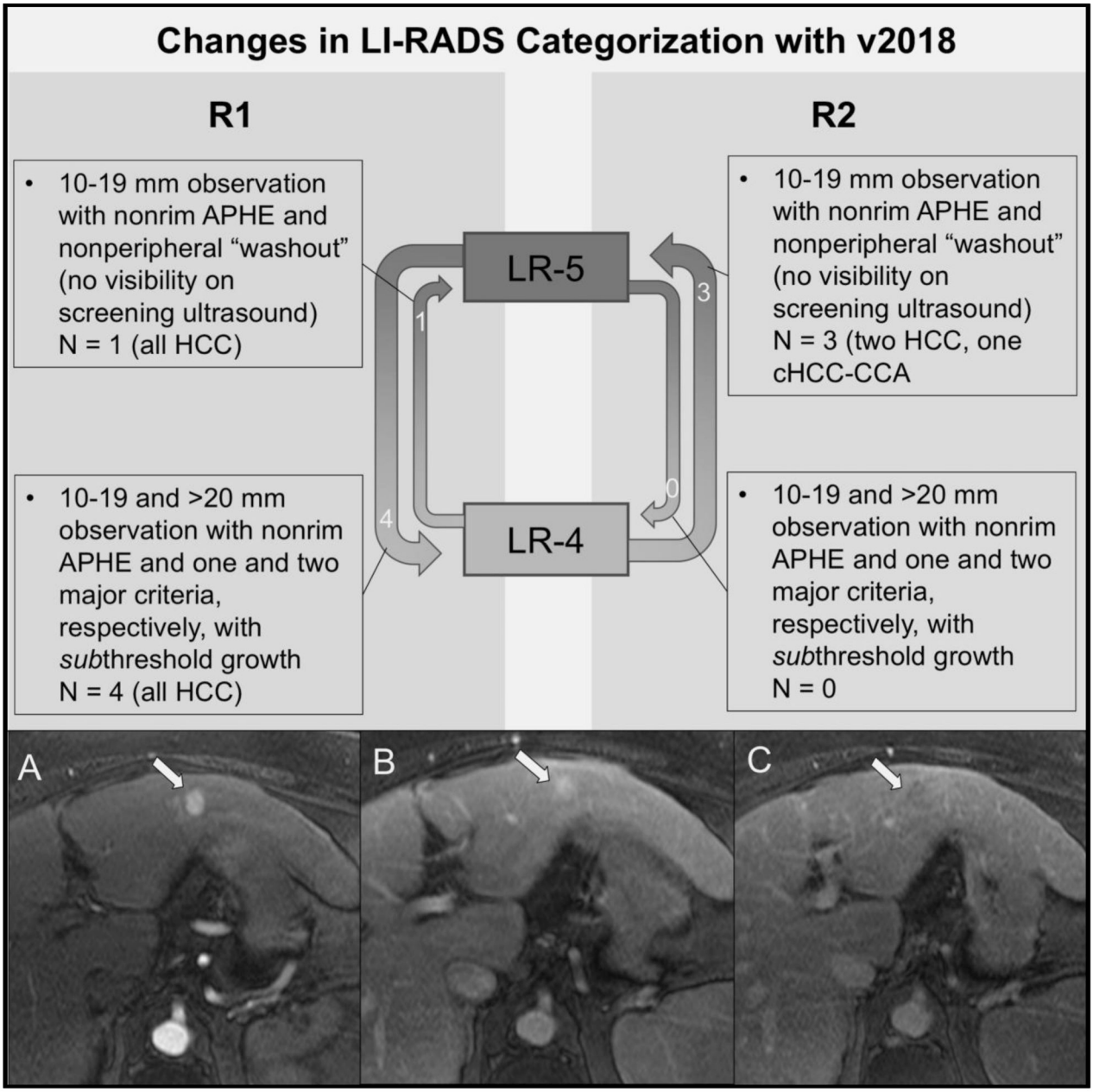

Fig 9.

Change in LI-RADS categorization with LI-RADS v2018. With v2018, observations were recategorized from LR-4 to LR-5 if they were 10-19 mm and demonstrated nonrim APHE and nonperipheral “washout”, but not have an enhancing “capsule”, meet criteria for threshold growth, or demonstrate visibility on antecedent ultrasound (one and three observations for R1 and R2, respectively). Observations were recategorized from LR-5 to LR-4 if they were 10-19 mm or ≥ 20 mm with one and two major criteria, respectively, and no longer met criteria for threshold growth (i.e., change between scans redefined as subthreshold growth per v2018; four and zero observations for R1 and R2, respectively). The bottom panel shows a HCC with different LI-RADS categories for v2017 versus v2018

for one reader. Patient is a 58-year-old woman with hepatitis C cirrhosis who underwent initial screening MR examination. Post-contrast arterial (A), portal venous (B), and delayed-phase (C) images after gadobenate dimeglumine (MultiHance) administration demonstrate an observation (arrows) at the junction of segments 2 and 3 with nonrim APHE and nonperipheral “washout”. R1 measured the lesion at 2.4 cm, corresponding to a score of LR-5 using v2017 and v2018. R2, on the other hand, measured the lesion at 1.9 cm and categorized it as LR-4 according to v2017, but the observation was scored LR-5 according to v2018.

Table 1:

Patient and Mass Characteristics

Patient Characteristics (n = 178)		
Gender	n (%)	*p* value *
Male	138 (78%)	…
Female	40 (22%)	<0.001
Age	Mean ± SD (range) in years	*p* value *
All	61.9 ± 8.4 (26-86)	…
Male	61.8 ± 8.1 (26-86)	…
Female	62.2 ± 9.4 (40-85)	0.80
Etiology of Cirrhosis	n (%)	*p* value *
Hepatitis C	83 (48%)	…
NASH	34 (20%)	…
Alcohol	35 (20%)	…
Cryptogenic	22 (12%)	…
Hepatitis B	8 (5%)	…
Other	16 (9%)	…
More than one factor	22 (13%)	…
Non-cirrhotic hepatitis B	4 (2%)	<0.001
Mass Characteristics (n = 178)		
Pathologic diagnosis	n (%)	*p* value *
Hepatocellular carcinoma (HCC)	105 (59%)	…
Combined hepatocellular-cholangiocarcinoma (cHCC-CCA)	48 (27%)	…
Intrahepatic cholangiocarcinoma (iCCA)	25 (14%)	<0.001
Source of tissue for pathologic diagnosis	n (%)	*p* value *
Hepatocellular carcinoma (N = 105)	…	…
Biopsy	0 (0%)	…
Resection	37 (35%)	…
Explant	68 (65%)	…
cHCC-CCA and iCCA (N = 73)	…	…
Biopsy	35 (48%)	…
Resection	22 (30%)	…
Explant	16 (22%)	<0.001
Interval from imaging to pathology	Mean ± SD in days	*p* value *
Hepatocellular carcinoma (n = 105)	189 ± 199	…
cHCC-CCA and iCCA (n = 73)	84 ± 132	<0.001
LRT between imaging and pathology	n (%)	*p* value *
Hepatocellular carcinoma (n = 105)	67 (64%)	…
cHCC-CCA and iCCA (n = 73)	15 (21%)	<0.001
Imaging modality for LI-RADS	n (%)	*p* value *
Hepatocellular carcinoma (n = 105)	…	…

Patient Characteristics (n = 178)		
MRI	89 (85%)	…
Extracellular GBCA	56 (63%)	…
Hepatobiliary GBCA	33 (37%)	…
CT	16 (15%)	…
cHCC-CCA and iCCA (n = 73)	…	…
MRI	60 (82%)	…
Extracellular GBCA	50 (83%)	…
Hepatobiliary GBCA	10 (17%)	…
CT	13 (18%)	0.80
Lesion size	Median (range)	*p* value **
Hepatocellular carcinoma (n = 105)	2.8 (1.4-19.0)	…
cHCC-CCA and iCCA (n = 73)	4.4 (1.5-14.0)	<0.001

* *p* values are based on results from Pearson χ^2 or Fisher exact test; $p < 0.005$ represents a significant difference.

** *p* value is based on the results from Mann-Whitney U test; $p < 0.005$ represents a significant difference.

Abbreviations: CT – computed tomography; GBCA – gadolinium based contrast agent; LI-RADS – Liver Imaging Reporting and Data System; LRT – locoregional therapy; MRI – magnetic resonance imaging; n – number; NASH – non-alcoholic steatohepatitis; SD – standard deviation

Table 2:

Diagnostic Performance of LI-RADS v2018 by Reader for Differentiating HCC from Non-HCC Malignancy

LI-RADS Category	Sensitivity* (%)	Specificity* (%)	PPV* (%)	NPV* (%)
LR-5 as a predictor of HCC	…	…	…	…
R1	65.7 (57.4-74.7)	89.0 (81.1-95.1)	89.6 (83.1-94.4)	64.4 (58.9-70.4)
R2	55.2 (46.7-65.0)	90.4 (82.7-96.1)	89.2 (81.8-94.5)	58.4 (53.8-63.8)
LR-5 or LR-TIV (definitely due to HCC) as a predictor for HCC	…	…	…	…
R1	67.6 (59.3-76.4)	84.9 (76.3-92.2)	86.6 (80.1-91.9)	64.6 (58.8-71.0)
R2	57.1 (48.6-66.8)	90.4 (82.7-95.9)	89.6 (82.4-94.6)	59.5 (54.7-64.9)
LR-M as a predictor for non-HCC	…	…	…	…
R1	65.8 (55.6-76.5)	89.5 (83.3-94.7)	81.4 (72.8-88.7)	79.0 (74.1-83.9)
R2	72.6 (62.7-82.4)	75.2 (67.3-83.1)	67.1 (60.1-74.5)	79.8 (74.0-85.4)
LR-M or LR-TIV (may be due to non-HCC malignancy) as a predictor for non-HCC	…	…	…	…
R1	76.7 (67.1-85.8)	89.5 (83.3-94.7)	83.6 (75.9-90.0)	84.7 (79.5-89.4)
R2	87.7 (79.5-94.2)	75.2 (67.3-83.1)	71.1 (64.8-77.6)	89.8 (83.9-94.2)
LR-M or LR-TIV (may be due to non-HCC malignancy) as a predictor for cHCC-CCA	…	…	…	…
R1	70.8 (58.2-83.0)	74.6 (67.6-81.8)	50.7 (43.5-59.3)	87.4 (82.6-91.6)
R2	62.3 (54.8-70.6)	85.4 (74.3-93.9)	45.6 (40.4-51.8)	92.0 (86.6-95.9)
LR-M or LR-TIV (may be due to non-HCC malignancy) as a predictor for iCCA	…	…	…	…
R1	88.0 (71.8-97.5)	70.5 (63.9-77.7)	32.8 (27.8-39.4)	97.3 (93.6-99.1)
R2	92.0 (76.9-99.0)	56.2 (39.2-64.2)	25.6 (22.3-29.8)	97.7 (93.3-99.4)

*Data in parentheses represent 95% confidence intervals

Abbreviations: cHCC-CCA – combined hepatocellular-cholangiocarcinoma; HCC – hepatocellular carcinoma; iCCA – intrahepatic cholangiocarcinoma; LI-RADS – Liver Imaging Reporting and Data System; LR-5 – definitely HCC; LR-M – probably or definitely malignant but not HCC specific; TIV – tumor in vein

Table 3:

LI-RADS Categories by Reader with Interrater Reliability Analysis

Overall Agreement by Category										
	LI-RADS Category for R2									
LI-RADS Category for R1	**LR-3**	**LR-4**	**LR-5**	**LR-M**	**LR-TIV (probably HCC)**	**LR-TIV (definitely HCC)**	**LR-TIV (maybe non-HCC)**	**All**	**κ Value***	**Agreement**
LR-3	1	1	0	0	0	0	0	2	…	…
LR-4	0	12	10	4	0	0	0	26	…	…
LR-5	1	5	50	20	0	1	0	77	…	…
LR-M	0	1	4	52	0	0	2	59	…	…
LR-TIV(probably HCC)	0	0	0	0	0	0	1	1	…	…
LR-TIV (definitely HCC)	0	0	1	0	0	1	3	5	…	…
LR-TIV (maybe non-HCC)	0	0	0	3	0	0	5	8	…	…
All	2	19	65	79	0	2	11	178	0.50 (0.40-0.60)	Moderate
Agreement on LR-5 or LR-TIV (definitely due to HCC) versus Other Categories										
	LI-RADS Category for R2									
LI-RADS Category for R1	**LR-5 or LR-TIV (definitely HCC)**					**Other**			**κ Value**	**Agreement**
LR-5 or LR-TIV (definitely HCC)	53					29			…	…
Other	14					82			0.51 (0.38-0.64)	Moderate
Agreement on LR-M or LR-TIV (may be due to non-HCC malignancy) versus Other Categories										
	LI-RADS Category for R2									
LI-RADS Category for R1	**LR-M or LR-TIV (maybe non-HCC)**					**Other**			**κ Value**	**Agreement**
LR-M or LR-TIV (not HCC)	62					5			…	…
Other	28					83			0.63 (0.52-0.74)	Substantial
Agreement on LR-5 versus LR-M or LR-TIV										
	LI-RADS Category for R2									
LI-RADS Category for R1	**LR-5**					**LR-M or LR-TIV (any)**			**κ Value**	**Agreement**
LR-5	50					21			…	…
LR-M or LR-TIV (any)	5					67			0.64 (0.51-0.76)	Substantial
Agreement on LR-5 or LR-TIV (definitely due to HCC) versus LR-M or LR-TIV (may be due to non-HCC malignancy)										
	LI-RADS Category for R2									

Abdom Radiol (NY). Author manuscript; available in PMC 2020 June 01.

Overall Agreement by Category										
	LI-RADS Category for R2									
LI-RADS Category for R1	**LR-3**	**LR-4**	**LR-5**	**LR-M**	**LR-TIV (probably HCC)**	**LR-TIV (definitely HCC)**	**LR-TIV (maybe non-HCC)**	**All**	**κ Value***	**Agreement**
LI-RADS Category for R1	**LR-5 or LR-TIV (definitely HCC)**					**LR-M or LR-TIV (maybe non-HCC)**			**κ Value**	**Agreement**
LR-5 or LR-TIV (definitely HCC)	52					23			…	…
LR-M or LR-TIV (maybe non-HCC)	4					60			0.62 (0.49-0.75)	Substantial

*Data in parentheses represent 95% confidence intervals

Abbreviations: HCC – hepatocellular carcinoma; LI-RADS – Liver Imaging Reporting and Data System; LR-3 – intermediate probability of malignancy; LR-4 – probably HCC; LR-5 – definitely HCC; LR-M – probably or definitely malignant but not specific for HCC specific; R1 – reader 1; R2 – reader 2; TIV – tumor in vein

Table 4:

Frequencies of Major Features by Reader among HCC versus non-HCC Malignancies with Interrater Reliability Analysis

Major Feature	HCC (n = 105)	Non-HCC Malignancy (n = 73)	*p* value*	Interpretation	κ Value†	Agreement
Nonrim APHE	…	…	…	…	0.60 (0.49-0.72)	Moderate
R1	91 (87%)	19 (26%)	<0.001	More common among HCC	…	…
R2	78 (74%)	10 (14%)	<0.001	More common among HCC	…	…
Nonperipheral “washout”	…	…	…	…	0.55 (0.42-0.67)	Moderate
R1	72 (69%)	16 (22%)	<0.001	More common among HCC	…	…
R2	65 (62%)	9 (12%)	<0.001	More common among HCC	…	…
Enhancing “capsule”	…	…	…	…	0.60 (0.48-0.72)	Moderate
R1	59 (56%)	15 (21%)	<0.001	More common among HCC	…	…
R2	51 (49%)	7 (10%)	<0.001	More common among HCC	…	…
Size	…	…	…	…	…	…
10-19 mm	…	…	…	…	0.48 (0.25-0.71)	Moderate
R1	11 (10%)	1 (1%)	0.04	No difference	…	…
R2	20 (19%)	6 (8%)	0.07	No difference	…	…
≥ 20 mm	…	…	…	…	0.48 (0.25-0.71)	Moderate
R1	94 (90%)	72 (99%)	0.04	No difference	…	…
R2	85 (81%)	67 (92%)	0.07	No difference	…	…
Threshold growth	…	…	…	…	−0.02 (−0.67-0.63)	Poor
R1	5 (5%)	2 (3%)	0.77	No difference		
R2	5 (5%)	2 (3%)	0.77	No difference		
Tumor in vein	…	…	…	…	0.67 (0.44-0.89)	Substantial
R1	2 (2%)	12 (16%)	0.001	More common among non-HCC		
R2	2 (2%)	10 (14%)	0.005	May be more common among non-HCC**		

*p values are based on results from Pearson χ^2 or Fisher exact test; $p < 0.005$ represents a significant difference.

**Falls on the borderline of significance.

†Data in parentheses represent 95% confidence intervals.

Abbreviations: HCC – hepatocellular carcinoma; APHE – arterial phase hyperenhancement; R1 – reader 1; R2 – reader 2

Table 5:

Frequencies of Criteria for LR-M by Reader among HCC versus non-HCC Malignancies with Interrater Reliability Analysis

Major Feature	HCC (n = 105)	Non-HCC Malignancy (n = 73)	*p* value*	Interpretation	K Value†	Agreement
Targetoid mass	…	…	…	…	0.61 (0.49-0.72)	Substantial
R1	9 (9%)	50 (68%)	<0.001	More common among non-HCC	…	…
R2	25 (24%)	63 (86%)	<0.001	More common among non-HCC	…	…
Rim APHE	…	…	…	…	0.55 (0.42-0.67)	Moderate
R1	5 (5%)	40 (55%)	<0.001	More common among non-HCC	…	…
R2	19 (18%)	57 (78%)	<0.001	More common among non-HCC	…	…
Peripheral “washout”	…	…	…	…	0.05 (−0.36-0.47)	Slight
R1	1 (1%)	13 (18%)	<0.001	More common among non-HCC	…	…
R2	3 (3%)	3 (4%)	0.97	No difference	…	…
Delayed central enhancement	…	…	…	…	0.54 (0.39-0.69)	Moderate
R1	5 (5%)	33 (45%)	<0.001	More common among non-HCC	…	…
R2	11 (10%)	37 (50%)	<0.001	More common among non-HCC	…	…
Targetoid restriction	…	…	…	…	0.34 (0.04-0.64)	Fair
R1	0 (0%)	5 (7%)	0.02	No difference	…	…
R2	4 (4%)	18 (25%)	<0.001	More common among non-HCC	…	…
Targetoid TP or HBP appearance	…	…	…	…	0.51 (0.16-0.87)	Moderate
R1	1 (1%)	6 (8%)	0.04	No difference	…	…
R2	3 (3%)	5 (7%)	0.37	No difference	…	…
Infiltrative appearance	…	…	…	…	0.24 (−0.08-0.55)	Fair
R1	2 (2%)	18 (25%)	<0.001	More common among non-HCC	…	…
R2	0 (0%)	8 (11%)	0.002	More common among non-HCC	…	…
Marked diffusion restriction	…	…	…	…	0.05 (−0.29-0.38)	Slight
R1	0 (0%)	2 (3%)	0.33	No difference	…	…
R2	6 (6%)	19 (26%)	0.002	More common among non-HCC	…	…
Necrosis or severe ischemia					0.07 (−0.36-0.51)	Slight
R1	0 (0%)	13 (18%)	<0.001	More common among non-HCC	…	…
R2	3 (3%)	2 (3%)	1.00	No difference	…	…

[*] p values are based on results from Pearson χ^2 or Fisher exact test; $p < 0.005$ represents a significant difference.

[†] Data in parentheses represent 95% confidence intervals.

Abbreviations: HCC – hepatocellular carcinoma; APHE – arterial phase hyperenhancement; HBP – hepatobiliary phase; LR-M – probably or definitely malignant but not HCC specific; R1 – reader 1; R2 – reader 2; TP – transitional phase

Abdom Radiol (NY). Author manuscript; available in PMC 2020 June 01.

Table 6:

Frequencies of Ancillary Features by Reader among HCC versus non-HCC Malignancies with Interrater Reliability Analysis

Major Feature	HCC (n = 105)	Non-HCC Malignancy (n = 73)	*p* value*	Interpretation	K Value†	Agreement
Favoring Malignancy, Not Specific for HCC						
US visibility as discrete nodule**	…	…	…	…	…	…
R1 / R2	21 (20%)	17 (23%)	0.73	No difference	…	…
Subthreshold growth	…	…	…	…	0.40 (0.21-0.60)	Fair
R1	24 (23%)	10 (14%)	0.18	No difference	…	…
R2	20 (19%)	6 (8%)	0.07	No difference	…	…
Corona enhancement	…	…	…	…	0.42 (0.10-0.74)	Moderate
R1	4 (4%)	5 (7%)	0.57	No difference	…	…
R2	6 (6%)	7 (10%)	0.49	No difference	…	…
Fat sparing in solid mass	…	…	…	…	0.20 (−0.37-0.78)	Slight
R1	1 (1%)	2 (3%)	0.75	No difference	…	…
R2	1 (1%)	5 (7%)	0.09	No difference	…	…
Restricted diffusion	…	…	…	…	0.41 (0.27-0.54)	Moderate
R1	17 (16%)	29 (40%)	0.007	No difference	…	…
R2	43 (41%)	43 (59%)	0.03	No difference	…	…
Mild-moderate T2 hyperintensity	…	…	…	…	0.56 (0.44-0.68)	Moderate
R1	36 (34%)	38 (52%)	0.03	No difference	…	…
R2	53 (50%)	45 (62%)	0.19	No difference	…	…
Iron sparing in solid mass	…	…	…	…	0.69 (0.42-0.96)	Substantial
R1	9 (10%)	0 (25%)	0.03	No difference	…	…
R2	8 (8%)	0 (11%)	0.04	No difference	…	…
Transitional phase hypointensity	…	…	…	…	0.33 (0.08-0.58)	Fair
R1	12 (11%)	2 (3%)	0.07	No difference	…	…
R2	16 (15%)	10 (14%)	0.94	No difference	…	…
Hepatobiliary phase hypointensity					0.59 (0.42-0.77)	Moderate
R1	21 (20%)	3 (4%)	0.005	Maybe more common among HCC	…	…
R2	19 (18%)	12 (16%)	0.93	No difference	…	…
Favoring HCC in Particular						
Nonenhancing “capsule”	…	…	…	…	0.00 (−0.47-0.47)	Slight
R1	14 (13%)	2 (3%)	0.03	No difference	…	…
R2	0 (0%)	0 (0%)	1.00	No difference	…	…
Nodule-in-nodule architecture	…	…	…	…	−0.03 (−0.68-0.63)	None
R1	4 (4%)	0 (0%)	0.24	No difference	…	…
R2	4 (4%)	1 (1%)	0.61	No difference	…	…
Mosaic architecture	…	…	…	…	0.47 (0.22-0.73)	Moderate
R1	17 (16%)	3 (4%)	0.02	No difference	…	…

Major Feature	HCC (n = 105)	Non-HCC Malignancy (n = 73)	*p* value*	Interpretation	K Value†	Agreement
R2	9 (9%)	2 (3%)	0.20	No difference	…	…
Fat in mass, more than adjacent liver	…	…	…	…	0.57 (0.34-0.79)	Moderate
R1	17 (16%)	4 (5%)	0.05	No difference	…	…
R2	10 (10%)	2 (3%)	0.14	No difference	…	…
Blood products in mass	…	…	…	…	0.53 (0.37-0.70)	Moderate
R1	21 (20%)	5 (7%)	0.03	No difference	…	…
R2	38 (36%)	4 (5%)	<0.001	More common among HCC	…	…

* *p* values are based on results from Pearson χ^2 or Fisher exact test; $p < 0.005$ represents a significant difference.

** Data provided to reader, therefore interrater reliability not assessed.

*** Falls on the borderline of significance

† Data in parentheses represent 95% confidence intervals.

Abbreviations: HCC – hepatocellular carcinoma; R1 – reader 1; R2 – reader 2; US – ultrasound