Title

Interreader Reliability of LI-RADS Version 2014 Algorithm and Imaging Features for Diagnosis of Hepatocellular Carcinoma: A Large International Multireader Study.

Journal

Radiology, 286(1)

ISSN

0033-8419

Authors

Fowler, Kathryn J
Tang, An
Santillan, Cynthia
et al.

Publication Date

2018

DOI

10.1148/radiol.2017170376

Peer reviewed

Radiology

Interreader Reliability of LI-RADS Version 2014 Algorithm and Imaging Features for Diagnosis of Hepatocellular Carcinoma: A Large International Multireader Study[1]

Kathryn J. Fowler, MD
An Tang, MD, MSc, FRCPC
Cynthia Santillan, MD
Mythreyi Bhargavan-Chatfield, PhD
Jay Heiken, MD
Reena C. Jha, MD
Jeffrey Weinreb, MD
Hero Hussain, MD
Donald G. Mitchell, MD
Mustafa R. Bashir, MD
Eduardo A. C. Costa, MD
Guilherme M. Cunha, MD
Laura Coombs, PhD
Tanya Wolfson, MA
Anthony C. Gamst, PhD
Giuseppe Brancatelli, MD
Benjamin Yeh, MD
Claude B. Sirlin, MD

[1] From the Mallinckrodt Institute of Radiology, Washington University School of Medicine, 510 S Kingshighway Blvd, St Louis, MO 63110 (K.J.F., J.H.); Department of Radiology, Centre Hospitalier de l'Universite de Montreal, Montreal, Canada (A.T.); Department of Radiology, Liver Imaging Group (C.S., C.B.S.), and Computational and Applied Statistics Laboratory, San Diego Supercomputer Center (T.W., A.C.G.), University of California San Diego, San Diego, Calif; American College of Radiology, Reston, Va (M.B., L.C.); Department of Radiology, MedStar Georgetown University Hospital, Washington, DC (R.C.J.); Department of Radiology, Yale Medical School, New Haven, Conn (J.W.); Department of Radiology, University of Michigan, Ann Arbor, Mich (H.H.); Department of Radiology, Thomas Jefferson University, Philadelphia, Pa (D.G.M.); Department of Radiology, Center for Advanced Magnetic Resonance Development, Duke University Medical Center, Durham, NC (M.R.B.); Cedrul, CT and MRI, Joao Pessoa, Brazil (E.A.C.C.); Clinica de Diagnostico por Imagem-CDPI-DASA, Rio de Janeiro, Brazil (G.M.C.); Division of Radiological Science, Di.Bi.Med., University of Palermo, Palermo, Italy (G.B.); and Department of Radiology, University of California San Francisco, San Francisco, Calif (B.Y.). Received February 15, 2017; revision requested April 10; revision received June 9; accepted July 18; final version accepted August 24. **Address correspondence to** K.J.F. (e-mail: *fowlerk@mir.wustl.edu*).

Supported by a research scholarship from the Fonds de Recherche du Québec–Santé and Fondation de l'association des radiologistes du Québec (FRQS-ARQ #26993).

Purpose: To determine in a large multicenter multireader setting the interreader reliability of Liver Imaging Reporting and Data System (LI-RADS) version 2014 categories, the major imaging features seen with computed tomography (CT) and magnetic resonance (MR) imaging, and the potential effect of reader demographics on agreement with a preselected nonconsecutive image set.

Materials and Methods: Institutional review board approval was obtained, and patient consent was waived for this retrospective study. Ten image sets, comprising 38–40 unique studies (equal number of CT and MR imaging studies, uniformly distributed LI-RADS categories), were randomly allocated to readers. Images were acquired in unenhanced and standard contrast material–enhanced phases, with observation diameter and growth data provided. Readers completed a demographic survey, assigned LI-RADS version 2014 categories, and assessed major features. Intraclass correlation coefficient (ICC) assessed with mixed-model regression analyses was the metric for interreader reliability of assigning categories and major features.

Results: A total of 113 readers evaluated 380 image sets. ICC of final LI-RADS category assignment was 0.67 (95% confidence interval [CI]: 0.61, 0.71) for CT and 0.73 (95% CI: 0.68, 0.77) for MR imaging. ICC was 0.87 (95% CI: 0.84, 0.90) for arterial phase hyperenhancement, 0.85 (95% CI: 0.81, 0.88) for washout appearance, and 0.84 (95% CI: 0.80, 0.87) for capsule appearance. ICC was not significantly affected by liver expertise, LI-RADS familiarity, or years of postresidency practice (ICC range, 0.69–0.70; ICC difference, 0.003–0.01 [95% CI: −0.003 to −0.01, 0.004–0.02]. ICC was borderline higher for private practice readers than for academic readers (ICC difference, 0.009; 95% CI: 0.000, 0.021).

Conclusion: ICC is good for final LI-RADS categorization and high for major feature characterization, with minimal reader demographic effect. Of note, our results using selected image sets from nonconsecutive examinations are not necessarily comparable with those of prior studies that used consecutive examination series.

Established hepatocellular carcinoma (HCC) imaging features include arterial phase hyperenhancement (APHE), washout appearance, and capsule appearance. Combinations of these features with size and growth are integral to imaging algorithms (1–6), as summarized in Table 1 (7,8). Unlike other algorithms, LI-RADS provides ordinal observation categories, a standardized lexicon, and major and ancillary features of HCC (Table 2).

Although there is extensive evaluation of imaging accuracy for HCC in the literature (9), the lack of a standardized lexicon, inconsistent and ambiguous definitions for imaging features, and variable protocols challenge the synthesis of available evidence. LI-RADS promotes standardization and, consequently, reproducibility across institutions and radiologists. Few studies have evaluated LI-RADS reproducibility, each with limitations of single-center retrospective cohorts and, often, with narrow focus on binary diagnosis of HCC rather than on the whole algorithm (10–12). The interreader reliability (IRR) of individual features and their incorporation into a complex algorithm remain gaps in knowledge.

The primary aim of our study is to determine in a large multicenter multireader setting the IRR of LI-RADS categories and major imaging features seen with CT and MR imaging and to determine the potential effect of reader demographics on agreement. The secondary aim is to assess the contribution of ancillary features and tie-breaking rules toward categorization.

Advances in Knowledge

- In this study of selected examinations and images showing liver lesions, the overall intraclass correlation coefficient (ICC) of Liver Imaging Reporting and Data System (LI-RADS) categorization was good for CT (0.67) and MR imaging (0.73), and the inclusion or exclusion of LI-RADS category M studies and the combination of LI-RADS categories 4 and 5 versus the separation of these categories did not affect overall agreement.
- Overall ICC was excellent for arterial phase hyperenhancement (0.87), washout appearance (0.85), and capsule appearance (0.84) and was similar for all major features at both CT and MR imaging.
- ICC was not significantly affected by liver imaging expertise, a priori LI-RADS familiarity, or years of postresidency practice (ICC range: 0.69–0.70; ICC difference, 0.003–0.01); there was borderline difference in agreement between readers in academic practice and those in a private or mixed practice environment (ICC difference, 0.009).
- Ancillary features and tie-breaking rules were used for 9.8% and 8.8% of review forms completed, respectively, and were more frequently used for MR imaging than for CT.
- Our results for ICC of major features and overall categorization were higher than those in previous publications; this may be related to the limitation of preselected observations rather than consecutive observations from clinical practice.

Implications for Patient Care

- LI-RADS interreader reliability is good for assigning the final category and excellent for characterizing major imaging features.
- LI-RADS interreader reliability was not significantly affected by a priori relative familiarity with LI-RADS classification, years of practice beyond residency, or expertise in liver imaging.

Materials and Methods

Collaboration and Logistical Support

Logistical support for this study was provided by the American College of Radiology (ACR). Two ACR employees (M.B., L.C.) are coauthors of this article. They helped design the study, oversaw the block randomizations and case assignments, and maintained the online case review forms and reader demographic database. Coauthors who are not employees of or consultants for ACR (K.J.F., C.S., C.B.S.) had control over the inclusion of all data and information submitted for publication.

Design

Institutional review board approval was obtained, and patient consent was waived for this Health Insurance Portability and Accountability Act–compliant retrospective multicenter multireader study.

Liver CT and MR Imaging Atlas

An atlas of CT and MR imaging LI-RADS observations was constructed from multiphase images contributed by nine radiologists (J.H., C.S., D.G.M., J.W., H.H., B.Y., M.B., R.C.J., and C.B.S.; approximately 9–35 years of experience) on the LI-RADS Steering Committee. All images were acquired or reviewed for clinical care from 2008 to 2013 at the members' institutions. Imaging studies

https://doi.org/10.1148/radiol.2017170376

Content code: GI

Radiology 2018; 286:173–185

Abbreviations:
ACR = American College of Radiology
APHE = arterial phase hyperenhancement
CI = confidence interval
HCC = hepatocellular carcinoma
ICC = intraclass correlation coefficient
IRR = interrater reliability
LI-RADS = Liver Imaging Reporting and Data System
LMM = linear mixed model
MCMC = Markov Chain Monte Carlo

Author contributions:
Guarantors of integrity of entire study, K.J.F., C.B.S.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, K.J.F., A.T., H.H., D.G.M., M.R.B., E.A.C.C., G.M.C., C.B.S.; clinical studies, K.J.F., A.T., C.S., J.H., R.C.J., H.H., D.G.M., M.R.B., G.M.C., G.B., B.Y., C.B.S.; statistical analysis, M.B., M.R.B., L.C., T.W., A.C.G.; and manuscript editing, K.J.F., A.T., C.S., M.B., J.H., R.C.J., J.W., D.G.M., M.R.B., E.A.C.C., T.W., A.C.G., G.B., B.Y., C.B.S.

Conflicts of interest are listed at the end of this article.

See also the article by Fraum et al in this issue.

Table 1

Diagnostic Imaging Criteria for Definite HCC

Diagnostic System	Imaging Criteria for Definite HCC	Specific Features
AASLD*	Lesion seen at antecedent US. At CT or MR imaging, lesion diameter >1 cm indicates APHE and washout	Applied in context of US surveillance.
OPTN†	At CT or MR imaging, for lesions ≥1 cm and <2 cm, one of the following: *(a)* APHE, washout, and capsule growth or *(b)* APHE and ≥50% growth in 6 months or less. For lesions ≥2 cm and ≤5 cm, one of the following: *(a)* APHE and washout, *(b)* APHE and capsule, or *(c)* APHE and threshold growth	Applied in context of MELD, exception point allocation for transplant candidates.
LI-RADS‡	For lesions 1.0 to 1.9 cm, APHE and two of the following: *(a)* washout, capsule growth, or threshold growth; *(b)* lesion seen at antecedent US and APHE and washout; or *(c)* APHE and ≥50% growth in 6 months or less. For lesions ≥2 cm, APHE and at least one of the following: washout, capsule growth, or threshold growth	Applied in all patients at risk. Additional considerations include ancillary features and tie-breaking rules§ for final categorization (LR-1 to LR-5, LR-5V, LR-M)

Note.—AASLD = American Association for the Study of Liver Disease, CT = computed tomography, LI-RADS = Liver Imaging Reporting and Data System, MELD, = Model for End-Stage Liver Disease, MR = magnetic resonance, OPTN = Organ Procurement and Transplantation Network, US = ultrasonography.

* Source.—Reference 1.

† Source.—Reference 7.

‡ Source.—Reference 8.

§ For further details and definitions regarding ancillary features and tie-breaking rules, please refer to the American College of Radiology (ACR) website: *http://www.acr.org/Quality-Safety/Resources/LIRADS/*.

Table 2

LI-RADS Version 2014 Categories and Definitions

LI-RADS Category	Definition
LR-1	Definitely benign
LR-2	Probably benign
LR-3	Intermediate probability for HCC
LR-4	Probably HCC
LR-5	Definitely HCC
LR-5V	Definite tumor in vein
LR-M	Probably malignant, not specific for HCC

were selected to represent a mix of classic imaging features and more equivocal findings, with the intention of capturing a wide range of observations that reflect institutional practice. For MR imaging, only observations obtained by using an extracellular space intravenous contrast agent or gadobenate were included. The LI-RADS observations ranged from benign to malignant. For each observation, the contributing Steering Committee member captured JPEG images of one or more representative sections through each acquired phase or sequence, inserted the images in one or more PowerPoint (Microsoft, Redmond, Wash) slides using a standard format, labeled all phases of contrast enhancement and sequences, marked the observation electronically, and annotated each slide with the observation's current size and, if available, prior size.

CT images depicting observations comprised standard protocol unenhanced and contrast material–enhanced images obtained in late arterial, portal venous, and delayed phases. Some CT images were supplemented with reconstructed images in the coronal or sagittal plane. On images in which relative attenuation of the observation to liver may not have been clearly portrayed, attenuation (in Hounsfield units) for the observation at different phases of contrast enhancement was provided.

MR images depicting observations were obtained with standard liver protocol sequences (eg, T2-weighted imaging, diffusion-weighted imaging, unenhanced T1-weighted in- and opposed-phase imaging, and dynamic unenhanced and contrast-enhanced imaging in the late arterial, portal venous, and delayed phases, along with arterial subtraction imaging).

Image Sets

All observations (n = 382) within the atlas were reviewed in consensus by two members of the LI-RADS Steering Committee (K.J.F., C.S.; 3 and 10 years of experience, respectively). Each observation was assigned a LI-RADS version 2014 category in consensus. The range of observation categories in the atlas included 21 LR-1 lesions, 56 LR-2 lesions, 96 LR-3 lesions, 74 LR-4 lesions, 91 LR-5 lesions, 28 LR-5V lesions, and 16 LR-M lesions. All observations were then divided into 10 case sets, each comprising 38–40 individual observations selected via block randomization to represent an equal representation

Table 3

Reader Demographic Information

Demographic Information	No. of Readers (n = 113)
Year completed residency	
1970–1989	18 (16)
1990–2004	49 (43)
2005–2014	46 (41)
Body fellowship trained	
Yes	89 (79)
No	23 (21)
Practice type	
Academic	74 (65)
Combined	34 (30)
Private	5 (4)
Liver transplant center	
Yes	75 (66)
No	38 (34)
Geographic region	
Asia	15 (13)
Europe	20 (18)
North America	68 (60)
Oceana	4 (4)
South America	6 (5)
Expertise in liver imaging	
I am a specialist	99 (87)
I read some liver images	13 (12)
I read very few liver images	1 (1)
Systems I currently use in practice*	
AASLD	48 (42)
APASL	3 (3)
EASL	23 (20)
JSH	5 (4)
LI-RADS	40 (35)
OPTN	42 (37)
No formal system	20 (18)
LI-RADS familiarity	
Not familiar at all	9 (8)
Somewhat familiar	65 (58)
Very familiar	39 (35)

Note.—Data in parentheses are percentages. AASLD = American Association for the Study of Liver Disease, APASL = Asian Pacific Association for the Study of the Liver, EASL = European Association for the Study of the Liver, JSH = Japanese Society of Hepatology, OPTN = Organ Procurement and Transplantation Network.

* For imaging systems, the respondents were allowed to enter more than one option, hence the cumulative percentages do not equal 100%. Some questions may have been unanswered, resulting in fewer than 113 total responses.

of CT and MR imaging and an approximately even distribution of LR-1 through LR-5, LR-5V, and LR-M findings. The number of observations per case set was selected to represent a large sample of observations but not more than could be reasonably completed in one review session. The finalized case sets were displayed in a standardized format and were subsequently converted from PowerPoint to Portable Document Format for distribution to readers.

Case Review Forms

Online case review forms were developed *(https://surveymonkey.com)*. Readers were required to enter the case identification number and, as quality assurance indicators, the modality (CT or MR imaging) and observation size. The review forms detailed the LI-RADS version 2014 diagnostic algorithm. Readers were asked to score the presence of major features, assign an initial LI-RADS category, score the presence of LI-RADS ancillary features, apply tie-breaking rules, and assign a final LI-RADS category. Each reader also completed a demographic survey to record their familiarity with LI-RADS, the number of years they had been in practice, their expertise in liver imaging with CT and MR imaging, and their experience with other imaging algorithms. As none of the case sets included gadoxetate-enhanced MR images or any US images, the review forms did not address imaging features specific to hepatobiliary contrast agents or the LR-5US category.

Readers

A total of 302 radiologists were invited to participate via e-mail, and 113 agreed to participate. Potential readers were selected based on their expertise in liver imaging, their institutional affiliation with a Steering Committee member, or referral from other radiologists. The list of radiologists was meant to capture broad representation of experts and nonexperts from many institutions in both academic and community settings, including those outside the United States. Radiologists were provided a brief introduction to the study and the rationale behind it. Radiologists who confirmed their participation received internet links to downloadable case sets, online review forms, instructions to complete the review forms, and case set assignment. Case sets were assigned by block randomization to these radiologists so that each case set would be reviewed by an equal number of readers. Readers were allowed access to LI-RADS version 2014 materials available online *(https://www.acr.org/Quality-Safety/Resources/LIRADS)* but received no additional training.

Statistical Analyses

Statistical analyses were performed with R software (version 3.3.1 [2016]; The R Foundation for Statistical Computing, Vienna, Austria). For Markov Chain Monte Carlo (MCMC) estimation, we used the rjags package *(https://*

cran.r-project.org/web/packages/rjags/rjags.pdf). Self-reported reader demographics were summarized. The use of ancillary features to up- or downgrade LI-RADS categories and the application of tie-breaking rules were summarized.

Mixed-effects regression analysis was used to estimate intraclass correlation coefficient (ICC), with 95% confidence intervals (CIs) as a measure of IRR for assigning LI-RADS categories and major features (13,14). ICC values were categorized according to Landis and Koch (15) as follows: ICC less than 0.20 indicated poor agreement; ICC of 0.21–0.40, fair agreement; ICC of 0.41–0.60, moderate agreement; ICC of 0.61–0.80, substantial agreement; ICC of 0.81–1.00, almost perfect agreement.

ICC accounts for the magnitude of disagreement between readers and, under certain conditions, is equal to the weighted κ value (16). A linear mixed model (LMM) was used to model LI-RADS categories as a function of cases and readers, with case- and reader-specific intercepts fitted. ICC was computed from the case, reader, and error variances of the

Table 4

Interrater Reliability for Assigning LI-RADS Categories and Major Features

Finding	CT	MR Imaging	Overall
LI-RADS category			
LR-1/LR-2, LR-3, LR-4, LR-5/LR-5V, LR-M	0.67 (0.61, 0.71)	0.73 (0.68, 0.77)	0.70 (0.67, 0.73)
LR-1/LR-2, LR-3, LR-4, LR-5/LR-5V	0.68 (0.62, 0.73)	0.73 (0.68, 0.77)	0.71 (0.67, 0.74)
LR-1/LR-2, LR-3, LR-4/LR-5/LR-5V, LR-M	0.61 (0.56, 0.66)	0.68 (0.63, 0.73)	0.65 (0.62, 0.69)
LR-1/LR-2, LR-3, LR-4/LR-5/LR-5V	0.63 (0.57, 0.68)	0.69 (0.64, 0.73)	0.67 (0.62, 0.70)
Major features			
APHE	0.86 (0.82, 0.91)	0.88 (0.83, 0.94)	0.87 (0.84, 0.90)
Washout appearance	0.84 (0.80, 0.90)	0.84 (0.80, 0.88)	0.85 (0.81, 0.88)
Capsule appearance	0.79 (0.72, 0.85)	0.88 (0.83, 0.92)	0.84 (0.80, 0.87)

Note.—Data are ICCs. Data in parentheses are 95% CIs. LR-M cases were included in major features analyses.

Figure 1

Figure 1: CT image set shows an LR-5 observation (arrows) with high agreement. The image set is exactly as it was presented to reviewers. All (100%) reviewers chose LR-5 as the category. *DP* = delayed phase *HAP* = hepatic arterial phase, *Pre* = precontrast, *PVP* = portal venous phase.

Figure 2

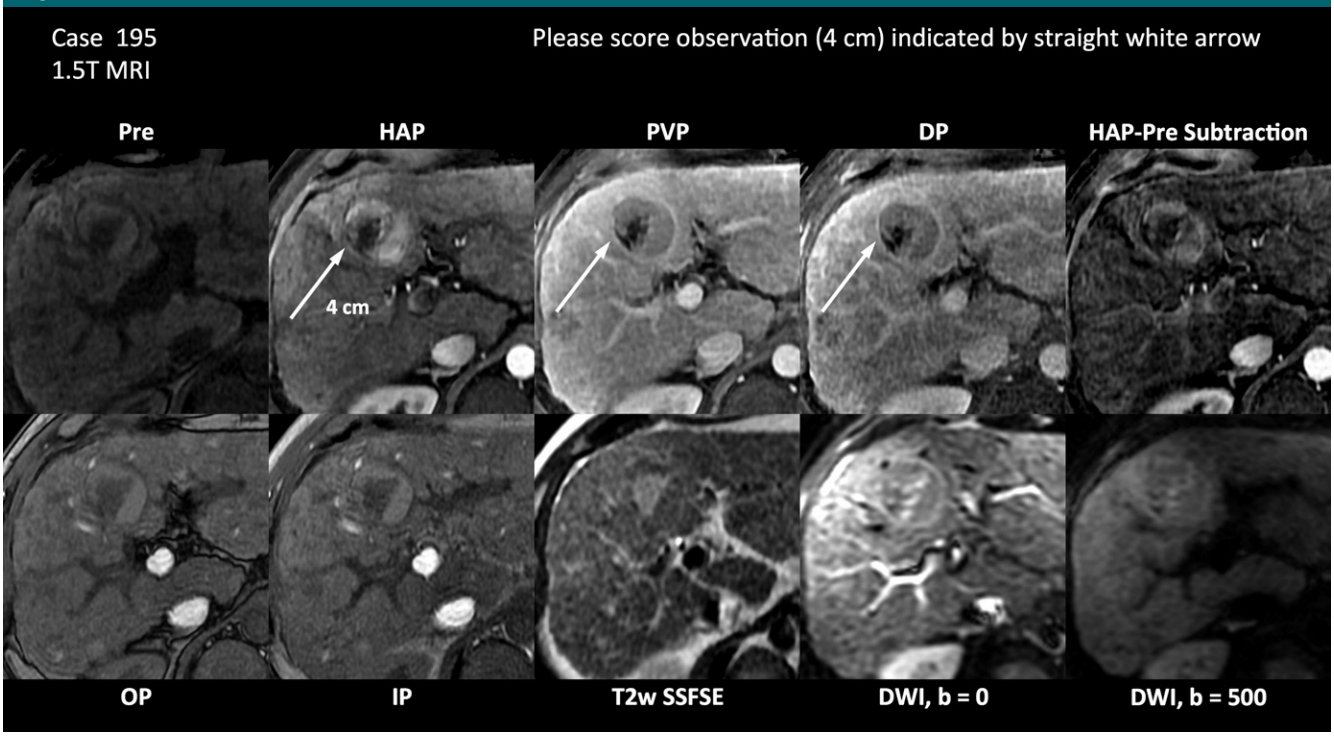

Figure 2: MR image set shows an LR-5 observation (arrows) with high agreement. The image set is exactly as it was presented to reviewers. All (100%) reviewers chose LR-5 as the category. *DP* = delayed phase, *DWI* = diffusion-weighted imaging, *HAP* = hepatic arterial phase, *IP* = in phase, *OP* = opposed phase, *Pre* = precontrast, *PVP* = portal venous phase, *SSFSE* = single-shot fast spin-echo, *T2W* = T2-weighted.

regression model. Parametric bootstrap analysis was used to construct 95% CIs around the ICCs. For IRR of LI-RADS categories, the following four stratifications were tested: *(a)* LR-1/LR-2, LR-3, LR-4, LR-5/LR-5 V/LR-M; *(b)* LR-1/LR-2, LR-3, LR-4, LR-5/LR-5V; *(c)* LR-1/LR-2, LR-3, LR-4/LR-5/LR-5V/LR-M; and *(d)* LR-1/LR-2, LR-3, LR-4/LR-5/LR-5V. In all stratifications, LR-1 and LR-2 were pooled because they represent observations that were probably or definitely benign and would require no follow-up (17). Similarly, LR-5 and LR-5V were pooled because they represent definite HCC and would require no biopsy according to current clinical practice guidelines. In two of the stratifications, LR-4 was pooled with LR-5 and LR-5V to form a composite category of probably or definitely HCC. Finally, because LR-M does not follow the same ordinal probability as other categories, the ICC was calculated both with and without inclusion of LR-M cases.

Similarly, a generalized LMM with case- and reader-specific intercepts was used to compute ICC for IRR on major features (binary outcomes). For these models, the estimation method was MCMC (18), and the 95% CIs were computed from the MCMC-generated distribution.

IRR was compared between participating radiologists and was based on their self-reported proficiency in liver imaging, familiarity with LI-RADS, years of posttraining practice, and overall expertise. The four characteristics were dichotomized (highly proficient vs other, very familiar with LI-RADS vs other, etc) and were examined in separate analyses. LMM and generalized LMMs were extended with the addition of reader group strata, which enabled computation of separate ICCs (eg, for experts vs nonexperts) within one model. The significance of the difference between reader subgroup ICCs was assessed by using parametric bootstrap analysis in the LMM and MCMC-generated distribution in the generalized LMM. LMM together with a parametric (model-based) bootstrap for 95% CI construction around the ICC and ICC differences were used. The LMM was refit 1000 times by using pseudodata samples. ICCs and their differences were computed at each iteration. The 95% CIs around each ICC and their difference were constructed from the bootstrap replicates. Two-tailed $P < .05$ was considered indicative of a significant difference. For these analyses, CT and MR imaging were combined, LR-M was included, and only the LR-1/LR-2, LR-3, LR-4, LR-5/LR-5V, LR-M stratification was used to assess agreement on ordinal LI-RADS categories. The *P* values derived from the analysis for categories and major features were not adjusted for multiple

Figure 3

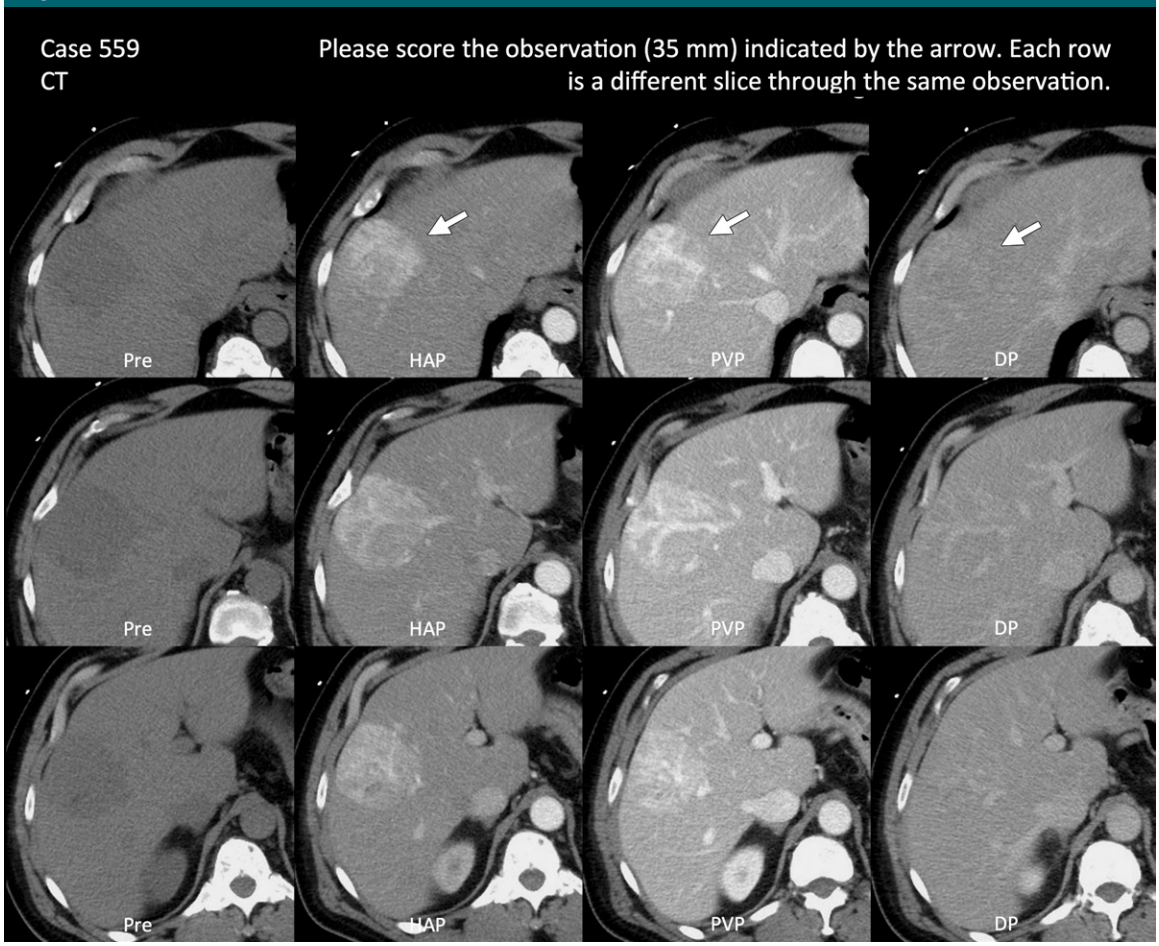

Figure 3: CT image set with low agreement. The image set is exactly as it was presented to reviewers. Consensus reading by the steering committee was LR-M. The most frequent category assigned by reviewers was LR-5, assigned by approximately 33% of reviewers. This image set is a challenging example, and the lesion (arrows) could arguably be categorized as LR-4. *DP* = delayed phase *HAP* = hepatic arterial phase, *Pre* = precontrast, *PVP* = portal venous phase.

comparisons to highlight any potential differences for further exploration.

Results

Readers

Of the 302 readers who were invited, 167 (55%) confirmed they would participate. Of these 167 who were sent demographic forms with links for case assignment, 54 (32%) dropped out, while the remaining 113 (68%) completed the case review forms. Table 3 summarizes demographics of the 113 final readers, who completed a total of 4346 case review forms. In instances in which duplicate case review forms were completed by the same reader, the second form was included in analysis under the assumption that the first form contained errors the reader intended to correct. Three cases and all corresponding review forms were excluded from final analysis because of incorrect labeling of contrast enhancement phases that was noticed only after case distribution. The final data set included 380 unique cases and 4009 separate review forms. Each case set was reviewed by an average of 11 readers (range, six to 17 readers). The unequal distribution of case sets was due to reader drop out and some readers downloading and scoring the wrong case sets. Readers were asked to input the case number for each form to ensure the appropriate case was scored.

Agreement on LI-RADS Categories

The agreement for LI-RADS categories was analyzed according to four distinct groups and is shown in Table 4. The ICC for agreement was highest for MR imaging for all groups, and agreement was slightly higher for variants without the combined LR-4/LR-5 category. The exclusion of LR-M had little effect on ICC. Figures 1–4 show examples of cases with high and

Figure 4

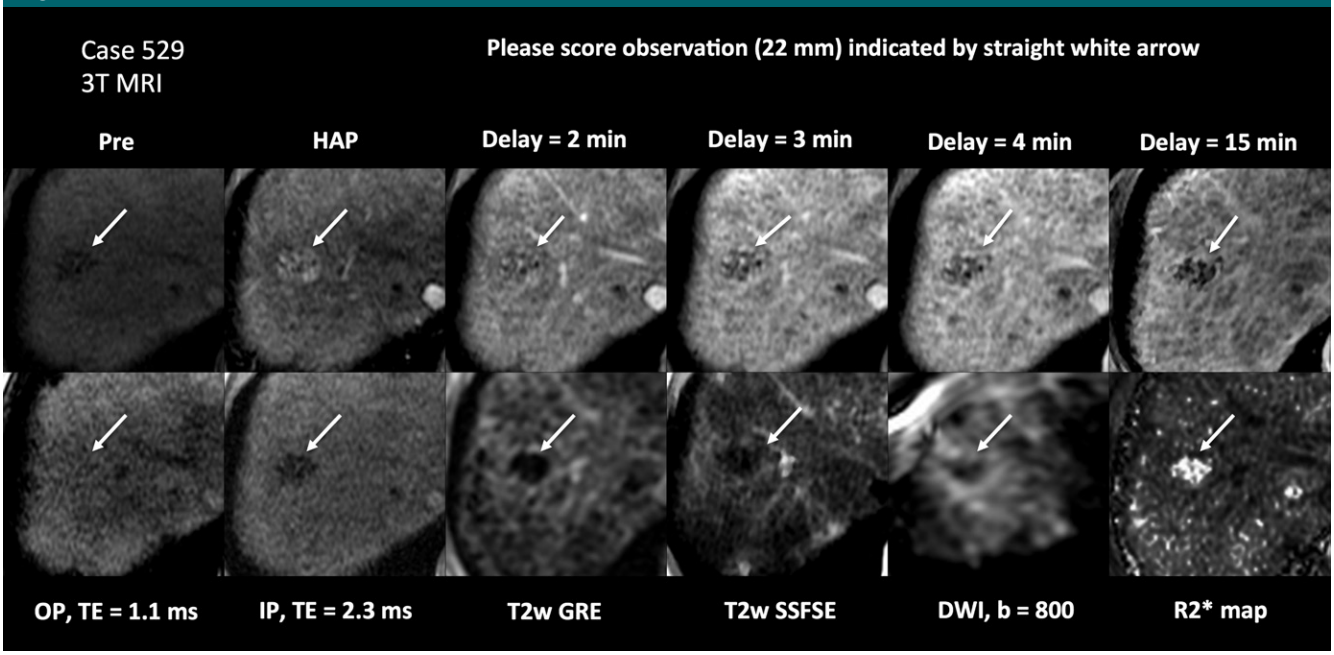

Figure 4: MR image set with low agreement. The image set is exactly as it was presented to reviewers. Although the most frequent response was LR-5, only 36% of reviewers scored this lesion (arrows) as LR-5. *DWI* = diffusion-weighted imaging, *GRE* = gradient-recalled echo, *HAP* = hepatic arterial phase, *IP* = in phase, *OP* = opposed phase, *Pre* = precontrast, *SSFSE* = single-shot fast spin-echo, *TE* = echo time, *T2W* = T2-weighted.

low agreement at both CT and MR imaging. Figures 5 and 6 are graphs of average score distribution for all cases at CT and MR imaging. Reader agreement was relatively high, with few outliers in each category and similar results between CT and MR imaging.

Agreement on Major Features

ICCs for assigning major features are shown in Table 4. Overall ICC was high for APHE (0.87; 95% CI: 0.84, 0.90), washout appearance (0.85; 95% CI: 0.81, 0.88), and capsule appearance (0.84; 95% CI: 0.80, 0.87). ICC was slightly higher with MR imaging than with CT for all features. Agreement was not analyzed for diameter or growth, which were provided to the readers. On the basis of the majority opinion, 50 lesions were smaller than 10 mm (CT, n = 16; MR imaging, n = 34), 126 were 10–19 mm (CT, n = 72; MR imaging, n = 54), and 204 were 20 mm or larger (CT, n = 104; MR imaging, n = 100).

Agreement and Reader Demographics

ICC differences were borderline higher for community practice readers than for academic readers (ICC difference, 0.009; P = .05). However, ICCs were not significantly affected by liver expertise, LI-RADS familiarity, or years of postresidency practice. Table 5 shows the agreement for overall categorization and major features based on reader demographics. Reader experience and expertise significantly affected assessment of washout appearance but did not affect other major features.

Ancillary Features

Ancillary features were used in 10% (393 of 4009) of review forms to modify the final category. Figure 7 shows the relative frequency of ancillary features in cases that were upgraded (164 MR imaging cases, 118 CT cases) or downgraded (118 MR imaging cases, 73 CT cases) by one category. Ancillary features were present more commonly on MR images than on CT images (74% [291 of 393] vs 26% [102 of 393]). The individual effect of each ancillary feature is difficult to discern because some cases that were upgraded or downgraded had multiple features selected. Although review forms asked specifically if ancillary features were used to determine the final category, they did require specification of which ones were used in the case of multiple selections.

Tie-Breaking Rules

Tie-breaking rules were used to change the final category on 9% (352 of 4009) of review forms, resulting in 201 individual cases (119 MR imaging cases, 82 CT cases).

Discussion

LI-RADS showed good ICC (range, 0.70–0.71) for LR-1 through LR-5 and LR-M category observations, with similar reliability shown for CT and MR imaging. The exclusion of LR-M findings did not meaningfully affect

Figure 5

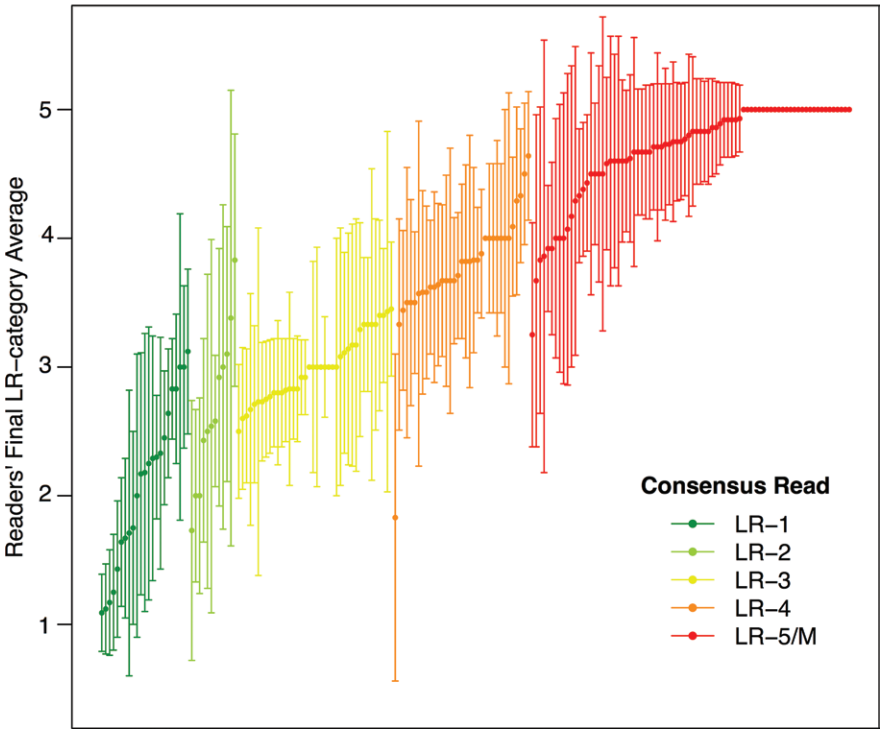

Figure 5: Graph shows reader agreement, with color coding of LI-RADS categories to show the distribution and average of scores for each individual image set at CT. *Consensus read* refers to the consensus reading of the steering committee. Each point represents the average score for a given observation. Bars represent the standard deviation.

overall reliability. ICC incorporates the magnitude of difference between readers, allowing for a global assessment of larger versus smaller disagreements between readers (ie, a disagreement between LR-1 and LR-5 categories has a greater effect on ICC than does a disagreement between LR-1 and LR-2 categories).

Six prior studies have assessed reader agreement regarding LI-RADS categories (11,19,20) or major features of HCC (11,19–22). Unlike these studies, which all included readers from the same institution, ours is further enhanced by *(a)* multicenter international cross-sectional reader pool, including community practice, academic, and mixed practice environments; *(b)* derivation of the cases from eight different sites; *(c)* the largest number of readers tested to date; *(d)* no training module aside from the LI-RADS materials available online; *(e)* mixture of all LI-RADS categories, including LR-M, at CT and MR imaging; *(f)* reduction of variability related to workstation adjustments (windowing, mislabeling of lesions, etc); *(g)* no requirement for pathologic proof of diagnosis, removing potential verification bias; and *(h)* thorough evaluation of LI-RADS, including data on use of ancillary features and tie-breaking rules. Table 6 compares the results of prior studies with ours.

The ICC for LI-RADS categories in our study (ie, 0.65–0.71) is similar to that reported in prior studies (ie, 0.44–0.82). We believe that our results accurately reflect reproducibility, given our large pool of cases (380 total, almost twice that of any other study) and our large pool of readers. These help control for problematic cases and equivocal features, which may negatively affect ICC.

Our major feature agreement was higher than that in previous publications (11,24). In particular, the ICC for capsule appearance in our study, 0.80 (95% CI: 0.74, 0.87) for CT and 0.89 (95% CI: 0.84, 0.94) for MR imaging, is higher than that reported in prior publications (0.37 and 0.65, respectively), with higher agreement for MR imaging in studies that evaluated both MR imaging and CT. Despite the intention to create an atlas with a mix of cases, the selection of cases may have favored those with more classic features and contributed to the higher ICC in our study. The lower agreement for "capsule" reported in the literature may be partly due to the difficulty in distinguishing a distinct delayed enhancing rim from the background fibrosis, which can be even more challenging in smaller lesions (25–27). Washout appearance agreement values were 0.85 (95% CI: 0.81, 0.88) for CT and 0.84 (95% CI: 0.79, 0.88) for MR imaging in our study; these values were greater than the 0.48–0.72 range reported in the literature. The subjective assessment of washout appearance may be confounded by background liver alterations (28), and some authors have suggested a quantitative approach to defining washout (29). The higher ICC results for major features in our study may underscore the variability that can be generated from workstation adjustments, which were removed in our study design. This raises the possibility that standardizing workstation adjustments may be a method to improve radiologist agreement, and this topic merits further investigation. Likewise, the selection of images potentially showing "classic" features may have contributed to higher ICC in our study. Additionally, the reviewers were forced to use the algorithm in a stepwise fashion, which may have improved reproducibility. Reader demographics minimally affected ICC, indicating LI-RADS can be applied reproducibly by readers of varying LI-RADS familiarity, liver expertise, or practice setting.

Ancillary features were frequently observed but resulted in a change in final category in only approximately

Figure 6

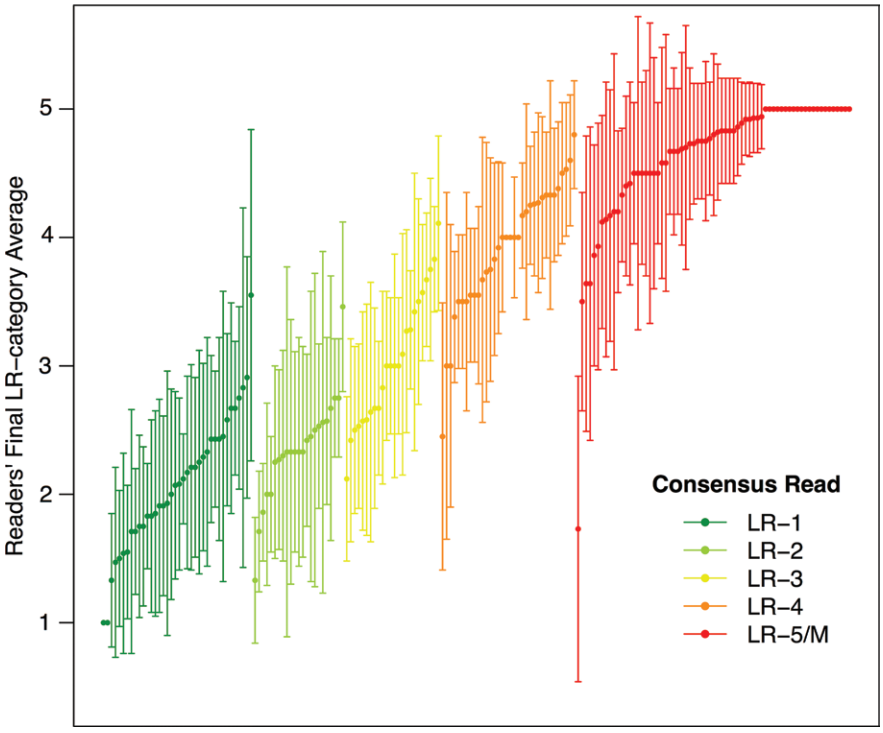

Figure 6: Graph shows reader agreement, with color coding of LI-RADS categories to show the distribution and average of scores for each individual image set at MR imaging. *Consensus read* refers to the consensus reading of the steering committee. Each point represents the average score for a given observation. Bars represent the standard deviation.

10% of total cases. Tie-breaking rules were used in a minority (9% of case review forms). Ancillary features and the use of tie-breaking rules were more prevalent for MR imaging, likely owing to a wider variety of image contrasts, the greater number of sequences, and the larger number of potential ancillary features available. The minor role of ancillary features in our study is similar to others' experience, suggesting that they may have a minimal effect on final diagnosis or reproducibility (30). The frequency and specificity of ancillary features requires further investigation in clinical trials to help determine their optimal incorporation in the imaging algorithm. The effect of individual ancillary features and tie-breaking rules on interreader agreement was not assessed with our methods, but it should be the topic of future study.

Our study had several limitations. To achieve a large interrater agreement study, case distribution was limited to prepared cases with select images and annotations. This construct may not be optimal for assessing all imaging features, many of which may require the reviewer to scroll or alter window settings to be

Table 5

Agreement on LI-RADS Category and Major Features by Reader Demographic Strata

Practice Type	Overall LI-RADS Category	APHE	Capsule	Washout Appearance
Academic	0.71 (0.67, 0.73)	0.86 (0.82, 0.90)	0.84 (0.80, 0.89)	0.85 (0.81, 0.89)
Community or mixed	0.69 (0.66, 0.72)	0.88 (0.82, 0.93)	0.83 (0.77, 0.89)	0.83 (0.77, 0.89)
P value	.050*	.613	.783	.680
Liver expertise				
Yes	0.70 (0.66, 0.73)	0.88 (0.84, 0.91)	0.83 (0.79, 0.87)	0.86 (0.82, 0.89)
Some or none	0.70 (0.66, 0.73)	0.81 (0.69, 0.89)	0.94 (0.74, 1.0)	0.75 (0.63, 0.85)
P value	.826	.134	.248	.0392*
LI-RADS familiarity				
Very familiar	0.70 (0.67, 0.74)	0.88 (0.81, 0.97)	0.83 (0.76, 0.89)	0.92 (0.86, 0.97)
Somewhat or not familiar	0.70 (0.67, 0.73)	0.87 (0.83, 0.90)	0.84 (0.79, 0.89)	0.82 (0.78, 0.86)
P value	.264	.820	.707	.0072*
Years of practice				
≥6 years	0.70 (0.66, 0.73)	0.86 (0.83, 0.90)	0.84 (0.80, 0.88)	0.84 (0.80, 0.87)
<6 years	0.71 (0.67, 0.74)	0.93 (0.83, 0.99)	0.85 (0.72, 0.98)	0.91 (0.81, 0.98)
P value	.852	.198	.911	.204

Note.—Bootstrap-based (for LI-RADS categories) and MCMC-based (for major features) *P* values for the differences in ICC between strata are shown. Data in parentheses are 95% CIs.

* *P* values indicate a significant difference. These *P* values are presented without an adjustment for multiple comparisons to highlight potential differences for further exploration.

Figure 7

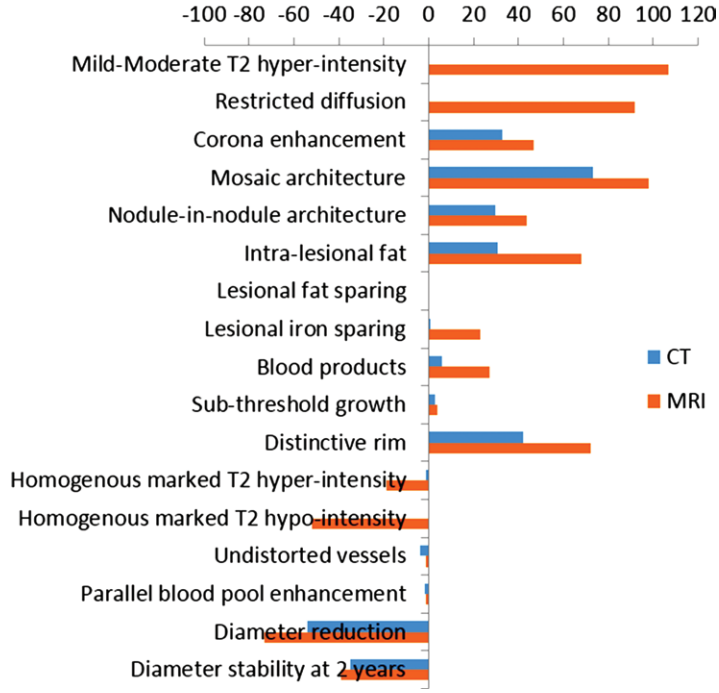

Figure 7: Relative frequency of ancillary features reported on review forms for cases that were up- or downgraded to achieve final category. Because some cases had multiple features selected, the effect of individual features on the final LI-RADS category is difficult to determine. The x-axis refers to the number of review forms marked for each feature.

confident. Of note, ambiguities related to feature designation may occur at different locations or section positions for each observation, and this may not have been captured in the select images included in the imaging atlas. Likewise, the observations included in the imaging atlas may represent higher quality data than encountered in routine practice and do not represent a consecutive case series; hence, our ICC values may not be duplicated in practice. For example, confident diagnosis of tumor in vein may require careful inspection of serial or reformatted images. Because such images were not routinely available, LR-5 and LR-5V were pooled. Thus, the prepared image set used in this study may be useful for training readers to use the LI-RADS system, but ICC obtained from this image set may not reflect interobserver agreement in real practice. Additionally, the very small number of consensus- or reader-identified LR-M tumors in the data made it unfeasible for us to treat LR-M as a separate category. More data would be required to answer the important question of separating different types of malignancy from nonmalignancy. As a compromise, we performed the analyses twice: once with LR-M pooled with LR-5 and LR-5V and once with consensus-identified LR-M excluded. An additional criticism of the current construct is the lack of a reference standard; thus, we did not assess diagnostic accuracy. Other studies have focused on assessment of the diagnostic accuracy of LI-RADS and other imaging algorithms, as validated with pathology (31,32). We did not assess ancillary features beyond the frequency of their use, nor did we include the LR-5US category.

In conclusion, in this multicenter study of interrater agreement on HCC imaging diagnosis, the overall agreement was good for final LI-RADS categorization and high for major feature characterization, with minimal reader demographic effects.

Our results for agreement for major features are higher than those in previous consecutive case series, possibly because variability from display setting adjustment was eliminated and because selected annotated images were assessed. Reader demographics showed little effect on ICC. Ancillary features and tie-breaking rules were used in a minority of cases and were more frequently used for MR imaging than for CT. Future studies are required to determine the optimal role of ancillary features for categorization.

Disclosures of Conflicts of Interest: K.J.F. disclosed no relevant relationships. **A.T.** disclosed no relevant relationships. **C.S.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: is a consultant for Robarts Clinical Research. Other relationships: disclosed no relevant relationships. **M.B.** disclosed no relevant relationships. **J.H.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: has patent agreements with Guerbet and Bayer. Other relationships: disclosed no relevant relationships. **R.C.J.** disclosed no relevant relationships. **J.W.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: is a consultant for Bracco, Guerbet, and Bayer Healthcare. Other relationships: disclosed no relevant relationships. **H.H.** disclosed no relevant relationships. **D.G.M.** disclosed no relevant relationships. **M.R.B.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: received grants from Siemens Healthcare, GE Healthcare, NGM Biopharmaceuticals, TaiwanJ Pharma, and Madrigal Pharmaceuticals. Other relationships: disclosed no relevant relationships. **E.A.C.C.** disclosed no relevant relationships. **G.M.C.** disclosed no relevant relationships. **L.C.** disclosed no relevant relationships. **T.W.** disclosed no relevant relationships. **A.G.** disclosed no relevant relationships. **G.B.** disclosed no relevant relationships. **B.Y.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: received a grant from GE Healthcare, is a shareholder in Nextrast, receives book royalties from Oxford University Press. Other relationships: disclosed no relevant relationships. **C.B.S.** disclosed no relevant relationships.

Table 6

Summary of Agreement from Literature

Study and LI-RADS Version	No. of Readers, No. of Lesions, and Case Selection Criteria	Modality and Format of Image Sets	Feature*	LI-RADS Category*	AF Use (%)
Davenport M (LI-RADS version 2013.1)†	10 readers (5 expert, 5 novice), 100 lesions, consecutive cases scored by research team from 2011–2013 to meet quota of 20 per LI-RADS category	MR imaging ECA phases, scrollable image sets	APHE, κ = 0.67 (95% CI: 0.65, 0.69); washout, κ = 0.48 (95% CI: 0.46, 0.5); capsule, κ = 0.52 (95% CI: 0.5, 0.54); diameter, ICC = 0.97 (95% CI: 0.96, 0.97)	LR-1, κ = 0.54 (95% CI: 0.51, 0.57); LR-2, κ = 0.11 (95% CI: 0.08, 0.14); LR-3, κ = 0.26 (95% CI: 0.23, 0.29); LR-4, κ = 0.28 (95% CI: 0.25, 0.31); LR-5, κ = 0.62 (95% CI: 0.59, 0.65)	22
Barth et al (LI-RADS version 2013.1)‡	Four readers (8–12 years of experience), 104 lesions, consecutive cases from database of LI-RADS categories screened for quota of about 15 per LI-RADS category	MR imaging with ECA, scrollable image sets, and a screenshot showing observation	APHE, κ = 0.51 (95% CI: 0.38, 0.65); washout, κ = 0.52 (95% CI: 0.39, 0.65); capsule, κ = 0.37 (95% CI: 0.23, 0.52); diameter, κ = 0.97 (95% CI: 0.965, 0.98)	All LI-RADS categories, κ = 0.44 (95% CI: 0.37, 0.52)	8
Zhang et al (LI-RADS version not specified)§	Two readers (6 years of postfellowship experience), 118 lesions, consecutive cases from existing database selected to include cases with both CT and MR images available	MR imaging with ECA, CT, and scrollable image sets	For CT: diameter, ICC = 0.95 (95% CI: 0.94, 0.97); APHE, ICC = 0.54 (95% CI: 0.44, 0.64); washout, ICC = 0.61 (95% CI: 0.46, 0.76); capsule, ICC = 0.56 (95% CI: 0.52, 0.6). For MR imaging: diameter, ICC = 0.98 (95% CI: 0.97, 0.99); APHE, ICC = 0.63 (95% CI: 0.54, 0.72); washout, ICC = 0.72 (95% CI: 0.64, 0.8); capsule, ICC = 0.65 (95% CI: 0.6, 0.7)	Not provided	NA
Ehman et al (LI-RADS version 2014)‖	Two readers (experience not specified), 184 lesions, consecutive cases with histopathology proof of diagnosis of HCC	MR imaging with ECA, CT, scrollable image sets	For CT: APHE, κ = 0.74; washout, κ = 0.64; capsule, κ = 0.46. For MR imaging: APHE, κ = 0.22; washout, κ = 0.58;capsule, κ = 0.62	ICC for all readers both CT and MR imaging.All LI-RADS, ICC = 0.59 (95% CI: 0.3, 0.95)	NA
Sofue et al (LI-RADS version 2013.1)#	Two readers (7 and 6 years of experience), 111 lesions, consecutive cases with hypervascular nodules by report	MR imaging with EC phases, scrollable image sets	Washout, κ = 0.75; capsule, κ = 0.70	Clinical and research reading; all LI-RADS categories, κ = 0.82	NA
Bashir et al (LI-RADS version 2013.1)**	Three readers (2, 4, and 12 years of experience), 200 lesions, consecutive cases with hypervascular nodules by report	MR imaging with ECA, CT, MR imaging with HBA, scrollable image sets	Washout, κ = 0.69 (95% CI: 0.61, 0.77); capsule, κ = 0.59 (95% CI: 0.51, 0.67); threshold growth, κ = 0.67 (95% CI: 0.59, 0.75); diameter, κ = 0.99 (95% CI: 0.99, 0.99)	LR-5 or <LR-5, κ = 0.68 (95% CI: 0.60, 0.76)	NA
Current study (LI-RADS version 2014)	113 readers (average 11 readers, diverse background), 380 lesions, selected cases from multiple institutions submitted to a liver atlas	MR imaging with ECA, CT, PowerPoint image captures	All CT: APHE, ICC = 0.86 (95% CI: 0.82,0.91); washout, ICC = 0.84 (95% CI: 0.80, 0.90); capsule, ICC = 0.80 (95% CI: 0.74, 0.87). All MR imaging: APHE, ICC = 0.91 (95% CI: 0.84, 0.96); washout, ICC = 0.83 (95% CI: 0.79, 0.88); capsule, ICC = 0.89 (95% CI: 0.84, 0.94).	ICC for all readers: MR imaging, 0.73 (95% CI: 0.68, 0.77); CT, 0.67 (95% CI: 0.61, 0.71)	9.8

Note.—A = arterial phase hyperenhancement, AF = ancillary features, C = capsule, D = diameter, ECA = extracellular agents, HBA = hepatobiliary agents, NA = not applicable, TG = threshold growth, W = washout.

* Data in parentheses are 95% CIs.

† Source.–Reference 11.

‡ Source.–Reference 19.

§ Source.–Reference 23.

‖ Source.–Reference 20.

Source.–Reference 21.

** Source.–Reference 22.

References

1. Bruix J, Sherman M; American Association for the Study of Liver Diseases. Management of hepatocellular carcinoma: an update. Hepatology 2011;53(3):1020–1022.
2. Wald C, Russo MW, Heimbach JK, Hussain HK, Pomfret EA, Bruix J. New OPTN/UNOS policy for liver transplant allocation: standardization of liver imaging, diagnosis, classification, and reporting of hepatocellular carcinoma. Radiology 2013;266(2):376–382.
3. Jha RC, Mitchell DG, Weinreb JC, et al. LI-RADS categorization of benign and likely benign findings in patients at risk of hepatocellular carcinoma: a pictorial atlas. AJR Am J Roentgenol 2014;203(1):W48–W69.
4. Korean Liver Cancer Study Group (KLCSG); National Cancer Center, Korea (NCC). 2014 Korean Liver Cancer Study Group-National Cancer Center Korea practice guideline for the management of hepatocellular carcinoma. Korean J Radiol 2015;16(3):465–522.
5. Kudo M, Matsui O, Izumi N, et al. Surveillance and diagnostic algorithm for hepatocellular carcinoma proposed by the Liver Cancer Study Group of Japan: 2014 update. Oncology 2014;87(Suppl 1):7–21.
6. Pomfret EA, Washburn K, Wald C, et al. Report of a national conference on liver allocation in patients with hepatocellular carcinoma in the United States. Liver Transpl 2010;16(3):262–278.
7. OPTN/UNOS policy 9: Allocation of Livers and Liver-Intestines. http://optn.transplant.hrsa.gov/ContentDocuments/OPTN_Policies.pdf-nameddest=Policy_09. Published 2015. Accessed April 27, 2015.
8. American College of Radiology. Liver Imaging Reporting and Data System version 2014. http://www.acr.org/Quality-Safety/Resources/LIRADS/. Accessed April 25, 2015.
9. Lee YJ, Lee JM, Lee JS, et al. Hepatocellular carcinoma: diagnostic performance of multidetector CT and MR imaging: a systematic review and meta-analysis. Radiology 2015;275(1):97–109.
10. Zhang YD, Zhu FP, Xu X, et al. Classifying CT/MR findings in patients with suspicion of hepatocellular carcinoma: comparison of liver imaging reporting and data system and criteria-free Likert scale reporting models. J Magn Reson Imaging 2016;43(2):373–383.
11. Davenport MS, Khalatbari S, Liu PS, et al. Repeatability of diagnostic features and scoring systems for hepatocellular carcinoma by using MR imaging. Radiology 2014;272(1):132–142.
12. Bashir MR, Korosec FR, Reeder SB. CME update: review articles and commentaries in JMRI. J Magn Reson Imaging 2014;40(4):778.
13. McGraw KO, Wong SP. Forming inferences about soe intraclass correlation coefficients. Psychol Methods 1996;1(1):30–46.
14. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol 2012;8(1):23–34.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–174.
16. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correaltion coefficient as measures of reliability. Educ Psychol Meas 1973;33(3):613–619.
17. Tanabe M, Kanki A, Wolfson T, et al. Imaging outcomes of Liver Imaging Reporting and Data System version 2014 category 2, 3, and 4 observations detected at CT and MR imaging. Radiology 2016;281(1):129–139.
18. Wu S, Crespi CM, Wong WK. Comparison of methods for estimating the intraclass correlation coefficient for binary responses in cancer prevention cluster randomized trials. Contemp Clin Trials 2012;33(5):869–880.
19. Barth BK, Donati OF, Fischer MA, et al. Reliability, validity, and reader acceptance of LI-RADS: an in-depth analysis. Acad Radiol 2016;23(9):1145–1153.
20. Ehman EC, Behr SC, Umetsu SE, et al. 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) 2016;41(5):963–969.
21. Sofue K, Sirlin CB, Allen BC, Nelson RC, Berg CL, Bashir MR. How reader perception of capsule affects interpretation of washout in hypervascular liver nodules in patients at risk for hepatocellular carcinoma. J Magn Reson Imaging 2016;43(6):1337–1345.
22. Bashir MR, Huang R, Mayes N, et al. Concordance of hypervascular liver nodule characterization between the organ procurement and transplant network and liver imaging reporting and data system classifications. J Magn Reson Imaging 2015;42(2):305–314.
23. Zhang YD, Zhu FP, Xu X, et al. Liver Imaging Reporting and Data System: substantial discordance between CT and MR for imaging classification of hepatic nodules. Acad Radiol, 2016;23(3):344–352.
24. Choi JY, Cho HC, Sun M, Kim HC, Sirlin CB. Indeterminate observations (liver imaging reporting and data system category 3) on MRI in the cirrhotic liver: fate and clinical implications. AJR Am J Roentgenol 2013;201(5):993–1001.
25. Yu JS, Lee JH, Chung JJ, Kim JH, Kim KW. Small hypervascular hepatocellular carcinoma: limited value of portal and delayed phases on dynamic magnetic resonance imaging. Acta Radiol 2008;49(7):735–743.
26. Kelekis NL, Semelka RC, Worawattanakul S, et al. Hepatocellular carcinoma in North America: a multiinstitutional study of appearance on T1-weighted, T2-weighted, and serial gadolinium-enhanced gradient-echo images. AJR Am J Roentgenol 1998;170(4):1005–1013.
27. Marrero JA, Hussain HK, Nghiem HV, Umar R, Fontana RJ, Lok AS. Improving the prediction of hepatocellular carcinoma in cirrhotic patients with an arterially-enhancing liver mass. Liver Transpl 2005;11(3):281–289.
28. Khan AS, Hussain HK, Johnson TD, Weadock WJ, Pelletier SJ, Marrero JA. Value of delayed hypointensity and delayed enhancing rim in magnetic resonance imaging diagnosis of small hepatocellular carcinoma in the cirrhotic liver. J Magn Reson Imaging 2010;32(2):360–366.
29. Liu YI, Shin LK, Jeffrey RB, Kamaya A. Quantitatively defining washout in hepatocellular carcinoma. AJR Am J Roentgenol 2013;200(1):84–89.
30. Rimola J, Forner A, Tremosini S, et al. Non-invasive diagnosis of hepatocellular carcinoma ≤ 2 cm in cirrhosis. diagnostic accuracy assessing fat, capsule and signal intensity at dynamic MRI. J Hepatol 2012;56(6):1317–1323.
31. Burke LM, Sofue K, Alagiyawanna M, et al. Natural history of liver imaging reporting and data system category 4 nodules in MRI. Abdom Radiol (NY) 2016;41(9):1758–1766.
32. Sofue K, Burke LM, Nilmini V, et al. Liver imaging reporting and data system category 4 observations in MRI: risk factors predicting upgrade to category 5. J Magn Reson Imaging 2017;46(3):783–792.