Title: A Multi-Center Assessment of Interreader Reliability of LI-RADS version 2018 for MRI and CT

Authors:

Cheng William Hong, MD MS[1,2], Victoria Chernyak MD MS[3], Jin-Young Choi MD[4], Sonia Lee MD[5], Chetan Potu BS[2], Timoteo Delgado BS[2], Tanya Wolfson MA[6], Anthony Gamst PhD[6], Jason Birnbaum MD[7], Rony Kampalath MD[5], Chandana Lall MD[8], James T Lee MD[9], Joseph W Owen MD[9], Diego A Aguirre MD[10], Mishal Mendiratta-Lala MD[11], Matthew S Davenport MD[11], William Masch MD[11], Alexandra Roudenko MD[12], Sara C Lewis MD[13], Andrea Siobhan Kierans MD[14], Elizabeth M Hecht MD[14], Mustafa R Bashir MD[15], Giuseppe Brancatelli MD[16], Michael L Douek MD[17], Michael A. Ohliger MD PhD[2], An Tang MD MSc[18], Milena Cerny MD[18], Alice Fung MD[19], Eduardo A Costa MD[20], Michael T Corwin MD[21], John P McGahan MD[21], Bobby Kalb MD[22], Khaled M Elsayes MD[23], Venkateswar R Surabhi MD[23], Katherine Blair MD[23], Robert M Marks MD[24], Natally M Horvat MD PhD[25], Shaun Best MD[26], Ryan Ash MD[26], Karthik Ganesan MD[27], Christopher R Kagay MD (deceased)[28], Avinash Kambadakone MD[29], Jin Wang MD[30], Irene Cruite MD[31], Bijan Bijan MD[32], Mark Goodwin MD[33], Guilherme M Cunha MD[34], Dorathy Tamayo-Murillo MD[2], Kathryn J Fowler MD[2], Claude B Sirlin MD[2]

Affiliations:

[1]Department of Radiology and Biomedical Imaging, University of California San Francisco, San Francisco, CA, USA

[2]Liver Imaging Group, Department of Radiology, University of California San Diego, San Diego, CA, USA

[3]Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, NY USA

[4]Department of Radiology, Yonsei University, South Korea

[5]Department of Radiology, University of California Irvine, Orange, CA, USA

[6]Computational and Applied Statistics Laboratory, University of California San Diego, San Diego, CA, USA

[7]Department of Radiology, New York University, New York, NY, USA

[8]Department of Radiology, University of Florida, Jacksonville, FL, USA

[9]Department of Radiology, University of Kentucky, Lexington, KY, USA
[10]Department of Radiology, Fundacion Santa Fe de Bogota, Colombia
[11]Department of Radiology, University of Michigan, Ann Arbor, MI, USA
[12]Department of Radiology, Allegheny Health Network, Pittsburgh, PA, USA
[13]Department of Radiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[14]Department of Radiology, New York-Presbyterian/Weill Cornell Medical Center, New York, NY, USA
[15]Departments of Radiology and Medicine, Duke University Medical Center, New York, NY, USA
[16]Section of Radiology, Department of Biomedicine, Neuroscience and Advanced Diagnostics (BiND), University Hospital "Paolo Giaccone", Palermo, Italy
[17]Department of Radiology, University of California Los Angeles, Los Angeles, CA, USA
[18]Department of Radiology, Radiation Oncology and Nuclear Medicine, Université de Montréal, Montréal, Québec, Canada
[19]Department of Radiology, Oregon Health & Science University, Portland, OR, USA
[20]CEDRUL - Centro de Diagnóstico por Imagem, João Pessoa, Brazil
[21]Department of Radiology, University of California Davis, Davis, CA, USA
[22]Radiology Limited, Tucson, AZ, USA
[23]Department of Abdominal Imaging, University of Texas MD Anderson Cancer Center, Houston, TX, USA
[24]Department of Radiology, Naval Medical Center San Diego, San Diego, CA, USA
[25]University of Sao Paulo / Hospital Sirio-Libanes, São Paulo, Brazil
[26]Department of Radiology, University of Kansas, Kansas City, KS, USA
[27]Sir.H.N.Reliance Foundation Hospital and Research Center, Mumbai, India
[28]Department of Radiology, California Pacific Medical Center, San Francisco, CA, USA
[29]Department of Radiology, Massachusetts General Hospital, Boston, MA, USA
[30]The 3rd Affiliated Hospital, Sun Yat-sen University, Guangzhou, China
[31]Inland Imaging, Spokane, WA, USA
[32]Sutter Medical Group, Sacramento, CA, USA
[33]Austin Health, Melbourne, Australia
[34]Department of Radiology, University of Washington, Seattle, WA

Corresponding author during submission and review:
Cheng William Hong, MD MS
University of California, San Francisco
Department of Radiology and Biomedical Imaging
513 Parnassus Ave, S255, Box 0628
San Francisco, CA 94143, USA

Corresponding author if accepted:
Claude B Sirlin, MD
University of California, San Diego
Liver Imaging Group, Department of Radiology
9500 Gilman Dr #0888
La Jolla, CA 92093, USA

Funding Information and Disclosures: This project is supported by a 2017 RSNA Resident Research Grant (RR1726). The authors also acknowledge grant support from National Institutes of Health T32 EB005970-09. An Tang was supported by a Clinical Research Scholarship – Senior Salary Award by the Fonds de recherche du Québec en Santé and Fondation de l'association des radiologistes du Québec (FRQS-ARQ #298509).

Manuscript Type: Original Research

Word Count for Text: 2741

Data Sharing Statement: Data generated or analyzed during the study are available from the corresponding author by request.

Acknowledgments: We gratefully acknowledge the contributions of the late Dr. Christopher Kagay towards the study design, data acquisition, and manuscript revisions.

A Multi-Center Assessment of Interreader Reliability of LI-RADS version 2018 for MRI and CT

Article Type: Original Research

Summary Statement: In an international multi-center reader study with scrollable case sets, overall moderate reader agreement was observed for the 2018 version of the Liver Imaging-Reporting and Data System.

Key Results:

1. In this retrospective study of 484 patients, the Liver Imaging-Reporting and Data System v2018 assessed using a modified four-category ordinal scale had moderate reader agreement (ICC, 0.68).
2. Binary agreement for probably or definitely malignant categories (ICC, 0.63) and LR-5 (ICC, 0.63) was moderate, whereas agreement for LR-M (ICC, 0.46) was poor.
3. Research-versus-research agreement differed from research-versus-clinical agreement (ICC, 0.68 vs. 0.63, *P*=.03), indicating differences between these environments that warrant further study.

List of Abbreviations:

HCC, hepatocellular carcinoma
ICC, intra-class correlation coefficient
LI-RADS, Liver Imaging-Reporting and Data System

ABSTRACT

Background: Various limitations have impacted research evaluating reader agreement for Liver Imaging-Reporting and Data System (LI-RADS).

Purpose: To assess reader agreement of LI-RADS in an international multi-center, multi-reader setting using scrollable images.

Materials and Methods: This retrospective study used de-identified clinical multiphase CT and MRI examinations and reports with at least one untreated observation from six institutions and three countries; only qualifying examinations were submitted. Examination dates were October 2017 – August 2018 at the coordinating center. One untreated observation per examination was randomly selected using observation identifiers, and its clinically assigned features were extracted from the report. The corresponding LI-RADS v2018 category was computed as a re-scored clinical read. Each examination was randomly assigned to two of 43 research readers who independently scored the observation. Agreement for an ordinal modified four-category LI-RADS scale (LR-1/2, LR-3, LR-4, LR-5/M/tumor in vein) was computed using intra-class correlation coefficients (ICC). Agreement was also computed for dichotomized malignancy (LR-4/LR-5/LR-M/LR-tumor in vein), LR-5, and LR-M. Agreement was compared between research-versus-research reads and research-versus-clinical reads.

Results: 484 patients (mean age, 62 years ±10 [SD]; 156 women; 93 CT, 391 MRI) were included. ICCs for ordinal LI-RADS, dichotomized malignancy, LR-5, and LR-M were 0.68 (95% CI: 0.62, 0.74), 0.63 (95% CI: 0.56, 0.71), 0.58 (95% CI: 0.50, 0.66), and 0.46 (95% CI: 0.31, 0.61) respectively. Research-versus-research reader agreement was higher than research-versus-clinical agreement for modified four-category LI-RADS (ICC, 0.68 vs. 0.62, $P = .03$) and for dichotomized malignancy (ICC, 0.63 vs. 0.53, $P = .005$), but not for LR-5 ($P = .14$) or LR-M ($P = .94$).

Conclusion: There was moderate agreement for Liver Imaging-Reporting and Data System v2018 overall. For some comparisons, research-versus-research reader agreement was higher than research-versus-clinical reader agreement, indicating differences between the clinical and research environments that warrant further study.

Introduction

Multiphase CT and MRI are instrumental in the noninvasive diagnosis and management of hepatic malignancies (1). The American College of Radiology Liver Imaging-Reporting and Data System (LI-RADS) standardizes the terminology, technique, interpretation, and reporting of liver imaging in at-risk patients (2,3). LI-RADS categorizes observations from LR-1 (definitely benign) to LR-5 (definitely hepatocellular carcinoma [HCC]) and also includes categories for malignant observations without characteristic HCC features.

Higher LI-RADS categories correspond to an increasing probability of HCC (4–10). In addition, higher LI-RADS categories also have an increasing probability of progression to HCC or other malignancy on follow-up imaging (11,12). The LR-5 category, intended to be diagnostic for HCC, has an estimated specificity of 89–99% (10,13–16). While determining accuracy is necessary, determination of precision including reader reliability is also necessary. Prior studies found moderate agreement for LI-RADS, but are limited by factors such as small single-center, single-modality image sets, and/or use of a small number of readers from a single center (16–24). A recent meta-analysis by Kang et al. found moderate agreement (κ, 0.70) for LI-RADS categorization, but only included MRI examinations, and 14 of the 15 studies had single-center readers (25). The only multi-center study by Fowler et al. had multiple contributing sites and many readers (26) and found moderate agreement (ICC, 0.67) for LI-RADS categorization; however, pre-selected image sets were used instead of fully scrollable examinations, which may overestimate reader agreement.

In addition, prior studies only assessed research reads, and no prior study has incorporated reads performed in a clinical environment. Research readers, being aware that their readings will be analyzed, may review cases and follow the LI-RADS algorithm more carefully. They can also read cases in a controlled environment with fewer distractions. However, research readers cannot access clinical information or prior imaging and reports, and they are unable to discuss cases with other radiologists or referring clinicians. Due to these factors, studies which focus exclusively on research

settings may not be fully generalizable to the clinical setting. Knowledge about LI-RADS performance in the clinical setting, including assessment of reliability or agreement with clinical reports, is needed but not yet available.

Thus, this study aims to assess reader agreement of LI-RADS in a large, international multi-center, multi-reader setting, while using scrollable imaging exams. This study also incorporates de-identified clinical reads to gain insight into reader agreement in the clinical setting.

Materials and Methods

Study Design

This was a retrospective multi-center, international reader study of clinically acquired multiphase LI-RADS CT and MRI examinations (**Figure 1**). The study was HIPAA-compliant and approved by local institutional review boards, with waiver of informed consent as the research was minimal risk. Six institutions from three countries (United States, South Korea, and Colombia) submitted de-identified examinations and reports from unique patients with at least one untreated observation to the coordinating center (UCSD). Examinations were uploaded to a cloud-based platform and assigned to two of 43 readers (**Supplemental Table 1**), randomized so that both readers were from separate institutions different from the submitting site, eliminating the possibility of familiarity bias. 20% of examinations were randomly selected to be read twice by one of the readers to assess intra-reader agreement.

Imaging Protocols

The modalities, contrast agents, and scanner vendors at each site are in **Supplemental Tables 2 and 3**.

Examination Selection

Submitting sites identified examinations that contained at least one untreated observation. Exclusion criteria were not applicable as only qualifying examinations were submitted. While there may have been minor differences in imaging protocols, all examinations adhered to LI-RADS technical recommendations and were reported in accordance with LI-RADS reporting requirements. The examination dates ranged from October 2017 – August 2018 at the coordinating center, dates from other sites were de-identified. The report provided for each reported observation a unique numeric identifier, a series/image number, an assigned category, and its major and ancillary features. For reports issued

clinically in a language other than English, the submitting radiologist translated the report into English.

Observation Selection

For each examination, an image analyst (C.P., 1 year of experience) at the coordinating center reviewed the de-identified report and selected one untreated observation using a random number generator to pick among the identifiers. The corresponding observation was electronically labeled with an arrow. This was done on the reported series and image number unless a different image identified the selected observation more clearly. The image analyst was a researcher with study-specific training for his tasks and software usage and was supervised by C.W.H. (radiology resident) and C.B.S. (>20 years of experience).

Research Reads

Labeled, de-identified examinations were uploaded to a cloud-based platform (Arterys). The platform provided standard capabilities including scrolling, panning, zooming, window-level adjustment, regions-of-interest, and measurement calipers. The readers were mostly subspecialty abdominal radiologists. Each reader scored the annotated observation using a standardized Research Electronic Data Capture form which included fields for the category and individual imaging features. Readers were provided a stepwise guide for the reading platform and case report forms. Readers could review reference materials, but no training was provided, reflecting clinical practice. The research reads were performed between May 2019 and October 2020.

Research Reader Questionnaire

Each reader completed a questionnaire on their geographic region, institutional affiliation, fellowship training, experience, familiarity with LI-RADS, and institutional practice patterns.

Clinical Reads

De-identified reports were parsed automatically by custom Python scripts (Python Software Foundation, Wilmington, DE), which extracted feature-level information. Reports were randomly selected for manual verification, and in cases where the clinical reports were incorrectly formatted, the features were extracted manually. The reported imaging features were used to recalculate the corresponding LI-RADS v2018 category while excluding features that require comparison to prior examinations as a re-scored "clinical read" (henceforth referred to as clinical reads). This was necessary as while most of the examinations were clinically reported using LI-RADS v2017, the research readers would apply LI-RADS v2018 and would not have access to prior examinations.

Statistical analysis

Data analysis was performed by statisticians T.W. and A.G. (both >25 years of experience) using R (R Foundation for Statistical Computing, Vienna, Austria). The research population and the reader questionnaire were summarized descriptively.

LI-RADS categories were combined into a four-category ordinal scale, with ascending risk of malignancy: LR-1/LR-2, LR-3, LR-4, and LR-5/LR-M (probably malignant, but not specific for HCC)/LR-tumor in vein. It was necessary to pool categories with low frequency (LR-1/LR-2) or which do not lend themselves to ordinal sub-ranking (LR-5/LR-M/LR-tumor in vein) to allow computation of overall agreement. Agreement was assessed using intra-class correlation coefficients (ICC). Generally, ICCs of <0.5, 0.5 – 0.75, 0.75 – 0.90, and >0.90 indicate poor, moderate, good, and excellent agreement respectively (27). Examination-level agreement was computed for research-versus-research reads and for research-versus-clinical reads. Sub-analyses were performed for MRI and CT.

Binary agreement was computed using ICCs for: 1) LI-RADS categories dichotomized as probably or definitely malignant (LR-4/5/M/TIV vs. LR-1/2/3); 2) LR-5; 3) LR-M; 4) major

features, and; 5) ancillary features present in ⩾5% of cases according to all reads (to ensure meaningful evaluation).

Non-parametric bootstrap with per-case resampling was used to compute 95% confidence intervals (CI) and to compare ICCs pairwise (28,29). As there are no previously published data on clinical reads, we considered the comparisons exploratory and did not correct for multiple statistical comparisons; *P* values < .05 were considered significant.

Results

Patient and Clinical Examination Characteristics

CT and MRI examinations from 484 unique patients were included (**Table 1**). Seventy-four (15%) examinations originated outside the United States. Patients included 156 (32%) women and 328 (68%) men, and they ranged from 21 to 95 years of age (mean age, 62 years ± 10 [SD]). Ninety-three (19%) examinations were performed with CT and 391 (81%) with MRI, the latter including 174 (36%) using extracellular agents and 217 (45%) using gadoxetic acid.

Research Reader Characteristics

The study included 43 research readers from 33 institutions and 9 countries (**Table 2**): 33 were from the United States, two from Canada, two from Brazil, and one each from China, South Korea, Colombia, India, Italy, and Australia. Forty-one readers (95%) reported fellowship training in abdominal imaging, the remaining two readers were current fellows in abdominal imaging. Thirty-nine readers (91%) self-identified themselves as experts in liver imaging. All readers mostly or almost exclusively read abdominal imaging in their daily clinical practice. Thirty-eight readers stated that their institution used LI-RADS in daily clinical practice (88%).

The readers reported an average of 11 years ± 6 [SD] of post-training radiology experience. Thirty-five readers (81%) were in an academic setting, 3 (7%) were in private practice, and 5 (12%) were in a hybrid practice setting.

Each research reader interpreted 15 – 32 examinations (mean ± SD, 23 ± 3 exams). Of those exams, on average, 18 examinations were MRI, and 4 examinations were CT. They reported spending 2 - 30 minutes (mean ± SD, 12 ± 5 minutes) per case.

Agreement for Modified Four-category LI-RADS Scale

The agreement for the modified scale is summarized in **Figure 2**. Agreement was moderate for MRI (ICC, 0.68, [95% CI: 0.61, 0.73]), CT (ICC, 0.68, [95% CI: 0.60, 0.74]), and both modalities combined (ICC, 0.68, [95% CI: 0.53, 0.80]). For all modalities, better reader agreement was observed between research-versus-research reads than between research-versus-clinical reads (ICC, 0.68, [95% CI: 0.61, 0.73] vs. 0.62, [95% CI: 0.56, 0.67], *P* = .03). Better reader agreement was also observed between research-versus-research reads for MRI (ICC, 0.68, [95% CI: 0.60, 0.74] vs. 0.61, [95% CI: 0.54, 0.67], *P* = .02) but not for CT (ICC, 0.68, [95% CI: 0.53, 0.80] vs. 0.66, [95% CI: 0.53, 0.76], *P* = .66).

Intra-reader agreement for the modified scale was better than inter-reader agreement between research reads (ICC, 0.84 [95% CI: 0.74 – 0.90] vs. 0.68, [95% CI: 0.61, 0.73], *P* = .002).

Figure 3 and **Supplemental Figure 1** show examples where readers disagreed and where readers agreed.

Agreement for Dichotomized LI-RADS Categories and for Individual Imaging Features

Agreement was moderate for dichotomized malignancy (LR-4/LR-5/LR-M/LR-TIV) (ICC, 0.63, [95% CI: 0.55, 0.70]) and moderate for LR-5 versus not (ICC, 0.58, [95% CI: 0.50, 0.66]) (**Figure 4**). Agreement for LR-M versus not was poor (ICC, 0.46, [95% CI: 0.31, 0.61]) (**Figure 4**). Better agreement for dichotomized malignancy was observed among research-versus-research reads compared to research-versus-clinical reads (ICC, 0.63, [95% CI: 0.55, 0.70] vs. 0.53, [95% CI: 0.46, 0.60], *P* = .005). No significant differences in agreement among research reads compared to research-versus-clinical reads were observed for LR-5 (ICC, 0.58, [95% CI: 0.50, 0.66] vs. 0.53, [95% CI: 0.47, 0.60], *P* =.14) or LR-M (ICC, 0.46, [95% CI: 0.31, 0.61] vs. 0.46, [95% CI: 0.32, 0.61], *P* = .94).

Agreement was moderate for major features including arterial phase hyperenhancement (ICC, 0.65, [95% CI 0.57, 0.72]), washout (ICC, 0.53, [95% CI: 0.46, 0.60]), and capsule (ICC, 0.50, [95% CI 0.42, 0.58]) (**Figure 5**). No difference in agreement was observed for research-versus-research reads compared to research-versus-clinical reads for arterial phase hyperenhancement (ICC, 0.65, [95% CI: 0.57, 0.72] vs. 0.61, [95% CI: 0.54, 0.67], *P* = .27), washout (ICC, 0.53, [95% CI: 0.46, 0.60] vs. 0.53, [95% CI: 0.46, 0.60], *P* = .93), or capsule (ICC, 0.50, [95% CI: 0.42, 0.58] vs. 0.47, [95% CI: 0.38, 0.54], *P* = .47).

For ancillary features, agreement was moderate for restricted diffusion (ICC, 0.50, [95% CI: 0.42, 0.59]) and for mild-moderate T2 hyperintensity (ICC, 0.58, [95% CI: 0.50, 0.66]) (**Figure 6**). Agreement was poor for transitional phase hypointensity (ICC, 0.16, [95% CI: 0.03, 0.30]) and hepatobiliary phase hypointensity (ICC, 0.44, [95% CI: 0.32, 0.55]). Better agreement among research reads than between research and clinical reads was observed for mild-moderate T2 hyperintensity (ICC, 0.58 [95% CI: 0.50, 0.66] vs 0.46 [95% CI: 0.38, 0.54], *P* = .01). No differences in reader agreement were observed for the other ancillary features.

Discussion

Prior studies assessing the reader agreement of Liver Imaging-Reporting and Data System (LI-RADS) have been limited by factors such as single-center nature, small number of readers, pre-selected images, and lack of comparison to clinical reads, and we performed a large multi-center, multi-reader study to begin addressing these gaps in knowledge. The overall inter-reader agreement for a modified four-category LI-RADS scale was moderate among research reads (ICC, 0.68) and when comparing re-scored clinical reads to research reads (ICC, 0.62). There was also moderate agreement for probably or definitely malignant categories (ICC, 0.63), for LR-5 (ICC, 0.58), and for all three major features (ICC, 0.50 – 0.65). For ancillary features, there was moderate agreement for restricted diffusion (ICC, 0.50) and mild-moderate T2 hyperintensity (ICC, 0.58), with poor agreement for transitional phase hypointensity (ICC, 0.16) and hepatobiliary phase hypointensity (ICC, 0.44).

A novel aspect of our study was the comparison between recomputed clinical reads and research reads. We found higher agreement between research reads than between research vs clinical reads for assignment of ordinal LI-RADS categories pooled over both modalities (ICC, 0.68 vs. 0.62, $P = .03$) and for MRI (ICC, 0.68 vs. 0.61, $P = .02$). Although this does not necessarily imply that agreement in the research environment will be higher than agreement in the clinical environment, these results indicate differences in interpretation between the clinical and research environment that warrant further study. One possibility is that although the clinical reads were generally performed by subspecialty abdominal radiologists, many of the research readers were from LI-RADS committees and self-identified themselves as experts in liver imaging. In the clinical setting, prior imaging and reports may result in anchoring bias towards prior categorizations (30–32).

Several prior studies have provided important insights (25,26,33–35). Fowler et al. found an ICC of LI-RADS category assignment of 0.67 overall, which is similar to our result of 0.68 (26). Their study however reported agreement of 0.84 – 0.87 for the major features,

which is higher than 0.50 – 0.65 in our study. This may be due to their use of selected image sets in comparison to our use of scrollable examinations, which may have showed the imaging features more clearly. A meta-analysis of fifteen studies by Kang et al. found a pooled κ of 0.66 – 0.72 for the major features, in comparison to the ICC of 0.50 – 0.65 in our study (25). These variations may be related to differences in study design, as they reported substantial study heterogeneity within the included studies. Similar to their study and other prior studies, our study found that non-rim arterial phase hyperenhancement had the highest reader agreement of the major features.

Our study had several limitations. First, although we did have international participation in this study, most of our cases and readers were from academic medical centers in North America. Additionally, 91% of our research readers were self-reported experts in liver imaging, and only 7% were in private practice. Thus, further evaluation of LI-RADS among community radiologists and medical centers outside of North America should be the focus of future work. Our study did not assess agreement of treatment response categories, and so our results only generalize to untreated observations. Annotating the observation may have introduced bias based on the selected image. We could only assess the reads recomputed using LI-RADS v2018 excluding features that depended on prior comparisons rather than the clinically reported categories, and the reader agreement for subthreshold and threshold growth could not be assessed. In addition, the number of possible pairs of research readers exceeded the number of examinations, which precluded meaningful evaluation of the effect of reader characteristics on agreement. Finally, we could not directly evaluate agreement between clinical reads. It is possible that clinical agreement is similar to research agreement, just that clinical reads are different from research reads.

In conclusion, Liver Imaging-Reporting and Data System (LI-RADS) v2018 generally has moderate agreement for observation categorization and feature characterization. Future research is needed to identify methods for reducing variability amongst readers, such as training, structured reporting, automated category computation based on reported features, or development of computer-aided categorization. In the meantime, it is

important to be mindful of this variability as it can substantially impact patient care, and selected patients should be referred to multidisciplinary tumor boards when feasible for consensus diagnostic and management decisions. At institutions without multidisciplinary tumor boards, double reading and/or referral of these patients to centers with such tumor boards should be considered. There are differences in interpretation between the research and clinical environments that warrant further study. Future research studies should also focus on the diagnostic performance of LI-RADS in the clinical setting, especially among community radiologists and in medical centers outside of North America, which remains an important knowledge gap in the validation of LI-RADS.

References

1. Bruix J, Sherman M. Management of hepatocellular carcinoma: An update. Hepatology. Wiley-Blackwell; 2011;53(3):1020–1022. doi: 10.1002/hep.24199.
2. Elsayes KM, Kielar AZ, Agrons MM, et al. Liver Imaging Reporting and Data System: an expert consensus statement. J Hepatocell carcinoma. Dove Press; 2017;4:29–39. doi: 10.2147/JHC.S125396.
3. Elsayes KM, Kielar AZ, Chernyak V, et al. LI-RADS: a conceptual and historical review from its beginning to its recent integration into AASLD clinical practice guidance. J Hepatocell carcinoma. 2019;6:49–69. doi: 10.2147/JHC.S186239.
4. Chen N, Motosugi U, Morisaka H, et al. Added Value of a Gadoxetic Acid-enhanced Hepatocyte-phase Image to the LI-RADS System for Diagnosing Hepatocellular Carcinoma. Magn Reson Med Sci. 2016;15(1):49–59. doi: 10.2463/mrms.2014-0149.
5. Choi SH, Byun JH, Kim SY, et al. Liver Imaging Reporting and Data System v2014 With Gadoxetate Disodium-Enhanced Magnetic Resonance Imaging: Validation of LI-RADS Category 4 and 5 Criteria. Invest Radiol. 2016;51(8):483–490. doi: 10.1097/RLI.0000000000000258.
6. Cha DI, Jang KM, Kim SH, Kang TW, Song KD. Liver Imaging Reporting and Data System on CT and gadoxetic acid-enhanced MRI with diffusion-weighted imaging. Eur Radiol. 2017;27(10):4394–4405. doi: 10.1007/s00330-017-4804-1.
7. Abd Alkhalik Basha M, Abd El Aziz El Sammak D, El Sammak AA. Diagnostic efficacy of the Liver Imaging-Reporting and Data System (LI-RADS) with CT imaging in categorising small nodules (10-20 mm) detected in the cirrhotic liver at screening ultrasound. Clin Radiol. 2017;72(10):901.e1-901.e11. doi: 10.1016/j.crad.2017.05.019.
8. Kim Y-Y, An C, Kim S, Kim M-J. Diagnostic accuracy of prospective application of the Liver Imaging Reporting and Data System (LI-RADS) in gadoxetate-enhanced MRI. Eur Radiol. 2018;28(5):2038–2046. doi: 10.1007/s00330-017-5188-y.
9. Liu W, Qin J, Guo R, et al. Accuracy of the diagnostic evaluation of hepatocellular carcinoma with LI-RADS. Acta Radiol. 2017;284185117716700. doi:

10.1177/0284185117716700.

10. CB van der P, CS L, CB S, et al. Accuracy of the Liver Imaging Reporting and Data System in Computed Tomography and Magnetic Resonance Image Analysis of Hepatocellular Carcinoma or Overall Malignancy-A Systematic Review. Gastroenterology. Gastroenterology; 2019;156(4):976–986. doi: 10.1053/J.GASTRO.2018.11.020.
11. 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. doi: 10.1148/radiol.2016152173.
12. Hong CW, Park CC, Mamidipalli A, et al. Longitudinal evolution of CT and MRI LI-RADS v2014 category 1, 2, 3, and 4 observations. Eur Radiol. Springer Verlag; 2019;29(9):5073–5081. doi: 10.1007/s00330-019-06058-2.
13. Cerny M, Bergeron C, Billiard J-S, et al. LI-RADS for MR Imaging Diagnosis of Hepatocellular Carcinoma: Performance of Major and Ancillary Features. https://doi.org/101148/radiol2018171678. Radiological Society of North America; 2018;288(1):118–128. doi: 10.1148/RADIOL.2018171678.
14. Kim Y-Y, Kim M-J, Kim EH, Roh YH, An C. Hepatocellular Carcinoma versus Other Hepatic Malignancy in Cirrhosis: Performance of LI-RADS Version 2018. Radiology. Radiological Society of North America Inc.; 2019;291(1):72–80. doi: 10.1148/radiol.2019181995.
15. Lee S, Kim S-S, Roh YH, Choi J-Y, Park M-S, Kim M-J. Diagnostic Performance of CT/MRI Liver Imaging Reporting and Data System v2017 for Hepatocellular Carcinoma: A Systematic Review and Meta-Analysis. Liver Int. John Wiley & Sons, Ltd; 2020;40(6):1488–1497. doi: 10.1111/LIV.14424.
16. Lee SM, Lee JM, Ahn SJ, Kang H-J, Yang HK, Yoon JH. LI-RADS Version 2017 versus Version 2018: Diagnosis of Hepatocellular Carcinoma on Gadoxetate Disodium–enhanced MRI. https://doi.org/101148/radiol2019182867. Radiological Society of North America; 2019;292(3):655–663. doi: 10.1148/RADIOL.2019182867.
17. Chen J, Kuang S, Zhang Y, et al. Increasing the sensitivity of LI-RADS v2018 for

diagnosis of small (10–19 mm) HCC on extracellular contrast-enhanced MRI. Abdom Radiol. Springer; 2021;46(4):1530–1542. doi: 10.1007/s00261-020-02790-2.

18. Zhang Y, Tang W, Xie S, et al. The role of lesion hypointensity on gadobenate dimeglumine–enhanced hepatobiliary phase MRI as an additional major imaging feature for HCC classification using LI-RADS v2018 criteria. Eur Radiol. Springer Science and Business Media Deutschland GmbH; 2021;31(10):7715–7724. doi: 10.1007/s00330-021-07807-y.
19. Cha DI, Choi GS, Kim YK, et al. Extracellular contrast-enhanced MRI with diffusion-weighted imaging for HCC diagnosis: prospective comparison with gadoxetic acid using LI-RADS. Eur Radiol. Springer; 2020;30(7):3723–3734. doi: 10.1007/s00330-020-06753-5.
20. Lee C min, Choi SH, Byun JH, et al. Combined computed tomography and magnetic resonance imaging improves diagnosis of hepatocellular carcinoma ≤ 3.0 cm. Hepatol Int. Springer; 2021;15(3):676–684. doi: 10.1007/s12072-021-10190-x.
21. Chung JW, Yu JS, Choi JM, Cho ES, Kim JH, Chung JJ. Subtraction images from portal venous phase gadoxetic acid-enhanced MRI for observing washout and enhancing capsule features in LI-RADS version 2018. Am J Roentgenol. American Roentgen Ray Society; 2020;214(1):72–80. doi: 10.2214/AJR.18.20797.
22. Hwang SH, Park S, Han K, Choi J young, Park YN, Park MS. Optimal lexicon of gadoxetic acid-enhanced magnetic resonance imaging for the diagnosis of hepatocellular carcinoma modified from LI-RADS. Abdom Radiol. Springer New York LLC; 2019;44(9):3078–3088. doi: 10.1007/s00261-019-02077-1.
23. Min JH, Kim JM, Kim YK, et al. A modified LI-RADS: targetoid tumors with enhancing capsule can be diagnosed as HCC instead of LR-M lesions. Eur Radiol. Springer Science and Business Media Deutschland GmbH; 2022;32(2):912–922. doi: 10.1007/s00330-021-08124-0.
24. Chen J, Zhou J, Kuang S, et al. Liver Imaging reporting and Data System category 5: MRI predictors of microvascular invasion and recurrence after hepatectomy for hepatocellular carcinoma. Am J Roentgenol. American Roentgen Ray Society;

2019;213(4):821–830. doi: 10.2214/AJR.19.21168.

25. Kang JH, Choi SH, Lee JS, et al. Interreader Agreement of Liver Imaging Reporting and Data System on MRI: A Systematic Review and Meta-Analysis. J. Magn. Reson. Imaging. John Wiley and Sons Inc.; 2020. doi: 10.1002/jmri.27065.
26. Fowler KJ, Tang A, Santillan C, et al. Interreader Reliability of LI-RADS Version 2014 Algorithm and Imaging Features for Diagnosis of Hepatocellular Carcinoma: A Large International Multireader Study. Radiology. Radiological Society of North America; 2018;286(1):173–185. doi: 10.1148/radiol.2017170376.
27. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. J Chiropr Med; 2016;15(2):155–163. doi: 10.1016/J.JCM.2016.02.012.
28. Streiner DL, Norman GR. Health Measurement Scales - A Practical Guide to Their Development and Use. . 2nd ed. New York: Oxford University Press, Inc; 1994.
29. Hinkley D V. Bootstrap Methods. J R Stat Soc Ser B. Wiley; 1988;50(3):321–337. doi: 10.1111/j.2517-6161.1988.tb01731.x.
30. Lee CS, Nagy PG, Weaver SJ, Newman-Toker DE. Cognitive and System Factors Contributing to Diagnostic Errors in Radiology. Am J Roentgenol. 2013;201(3):611–617. doi: 10.2214/AJR.12.10375.
31. Bruno MA, Walker EA, Abujudeh HH. Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction. RadioGraphics. Radiological Society of North America; 2015;35(6):1668–1676. doi: 10.1148/rg.2015150023.
32. Busby LP, Courtier JL, Glastonbury CM. Bias in Radiology: The How and Why of Misses and Misinterpretations. https://doi.org/101148/rg2018170107. Radiological Society of North America; 2017;38(1):236–247. doi: 10.1148/RG.2018170107.
33. Abdel Razek AAK, El-Serougy LG, Saleh GA, Abd El-Wahab R, Shabana W. Interobserver Agreement of Magnetic Resonance Imaging of Liver Imaging Reporting and Data System Version 2018. J Comput Assist Tomogr. Lippincott Williams and Wilkins; 2020;44(1):118–123. doi: 10.1097/RCT.0000000000000945.
34. Schellhaas B, Hammon M, Strobel D, et al. Interobserver and intermodality agreement of standardized algorithms for non-invasive diagnosis of hepatocellular

carcinoma in high-risk patients: CEUS-LI-RADS versus MRI-LI-RADS. Eur Radiol. 2018; doi: 10.1007/s00330-018-5379-1.

35. Lim K, Kwon H, Cho J. Inter-reader agreement and imaging-pathology correlation of the LI-RADS M on gadoxetic acid-enhanced magnetic resonance imaging: efforts to improve diagnostic performance. Abdom Radiol (New York). Springer; 2020; doi: 10.1007/s00261-020-02421-w.

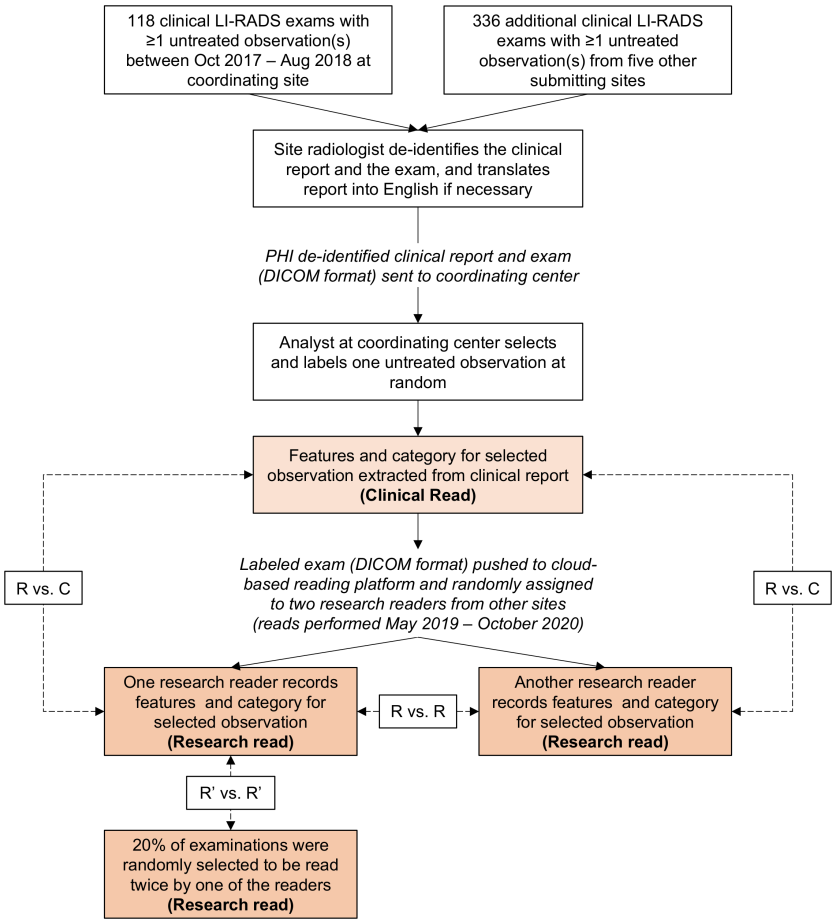


Figure 1: Schematic of the retrospective study design. De-identified examinations from the coordinating site and five other submitting sites were randomly assigned to two of 43 research readers for research reads. Features and category were extracted from the clinical report. This permitted the computation of inter-reader agreement between the research readers (R vs. R) and between the research and clinical readers (R vs. C). 20% of cases were also read twice by one of the research readers to permit the computation of intra-reader agreement (R' vs. R').

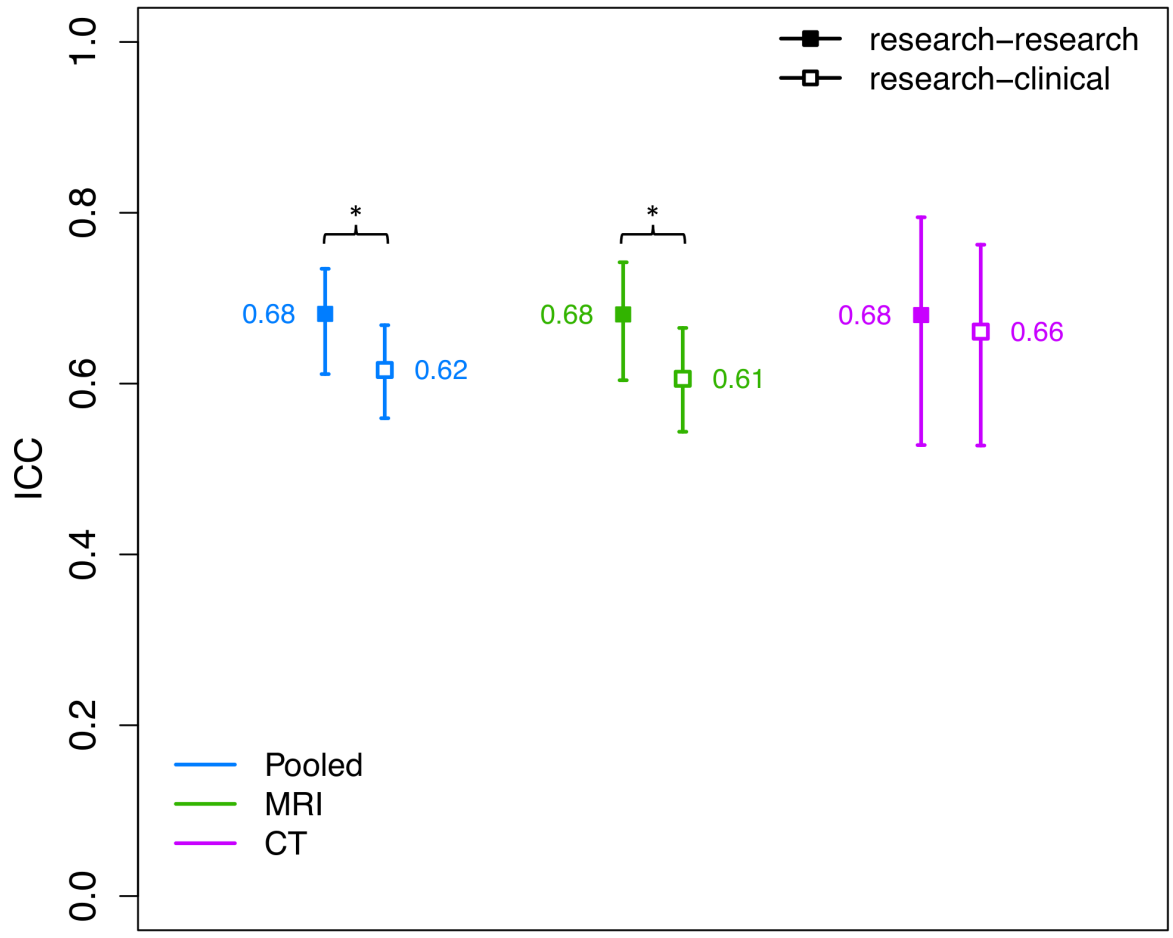


Figure 2. Reader agreement for modified four-category LI-RADS version 2018 scale based on imaging modality. Plot shows interclass correlation coefficients (ICC) for both modalities (blue), for MRI only (green), and for CT only (purple). Agreement among research reads only (filled squares) and between research and clinical reads (open squares) are shown. Tails represent 95% confidence intervals. *Represents a *P* value of < .05 by non-parametric bootstrap with per-case resampling. Research-versus-research agreement pooled over both modalities and for MRI only was better than research-versus-clinical agreement.

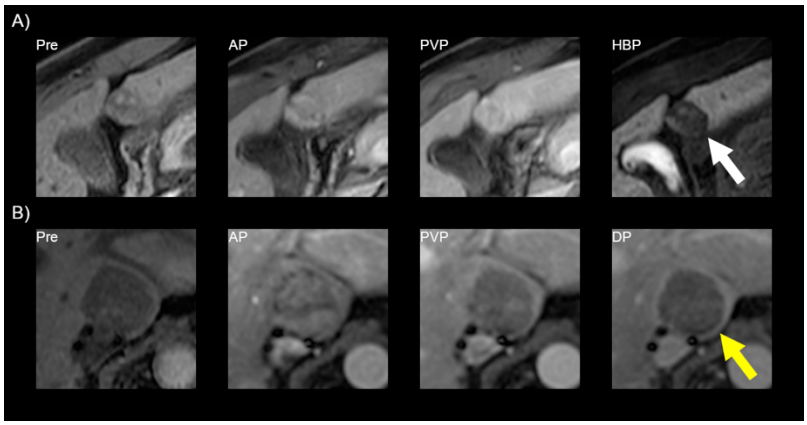


Figure 3. MR images showing examples where readers disagreed (A) and where readers agreed (B). The top row (A) shows images from an MR examination performed with gadoxetic acid from a 56-year-old male with cirrhosis secondary to hepatitis C. From left to right: pre-contrast (Pre), arterial phase (AP), portal venous phase (PVP), hepatobiliary phase (HBP). This 21 mm hepatobiliary phase hypointense observation (white arrow) was characterized on the clinical read as having non-rim arterial phase hyperenhancement and washout appearance and was categorized as LI-RADS category LR-5 (definitely hepatocellular carcinoma [HCC]). The first research reader characterized it as having a targetoid appearance and categorized it as LR-M (probably or definitely malignant, not specific for HCC). The second research reader characterized it as having no major features and paralleling the blood pool and categorized it as LR-2 (probably benign). It was subsequently resected and found to be a well-differentiated HCC. The bottom row (B) shows images from an MR examination performed with an extracellular contrast agent from a 61-year-old female with cirrhosis secondary to hepatitis C. From left to right: pre-contrast (Pre), arterial phase (AP), portal venous phase (PVP), delayed phase (DP). This 31 mm observation (yellow arrow) in the caudate lobe was characterized on the clinical read as having arterial phase hyperenhancement, washout appearance, and capsule appearance was categorized as LI-RADS category LR-5 (definitely hepatocellular carcinoma). Both research readers also categorized this observation as LR-5. The patient passed away from intracranial hemorrhage a few months later.

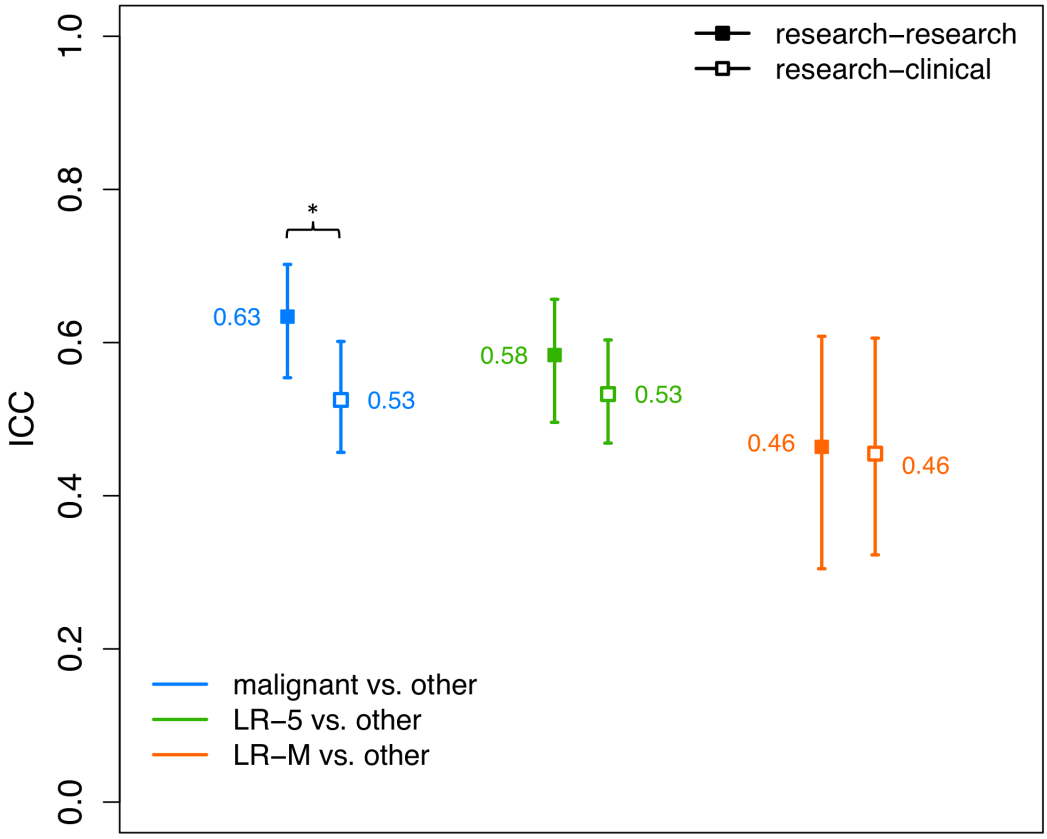


Figure 4. Reader agreement for dichotomized classification of LI-RADS version 2018. Plot shows interclass correlation coefficients (ICC) for the following dichotomized categories: probably or definitely malignant vs not (blue), LR-5 (definitely hepatocellular carcinoma [HCC]) vs. not LR-5 (green), and LR-M (probably or definitely malignant, not specific for HCC) vs not LR-M (orange). Agreement among research reads only (filled squares) and between research and clinical reads (open squares) are shown. Tails represent 95% confidence intervals. *Represents a *P* value of < .05 by non-parametric bootstrap with per-case resampling. Research-research agreement for malignant categories was better than research-clinical agreement.

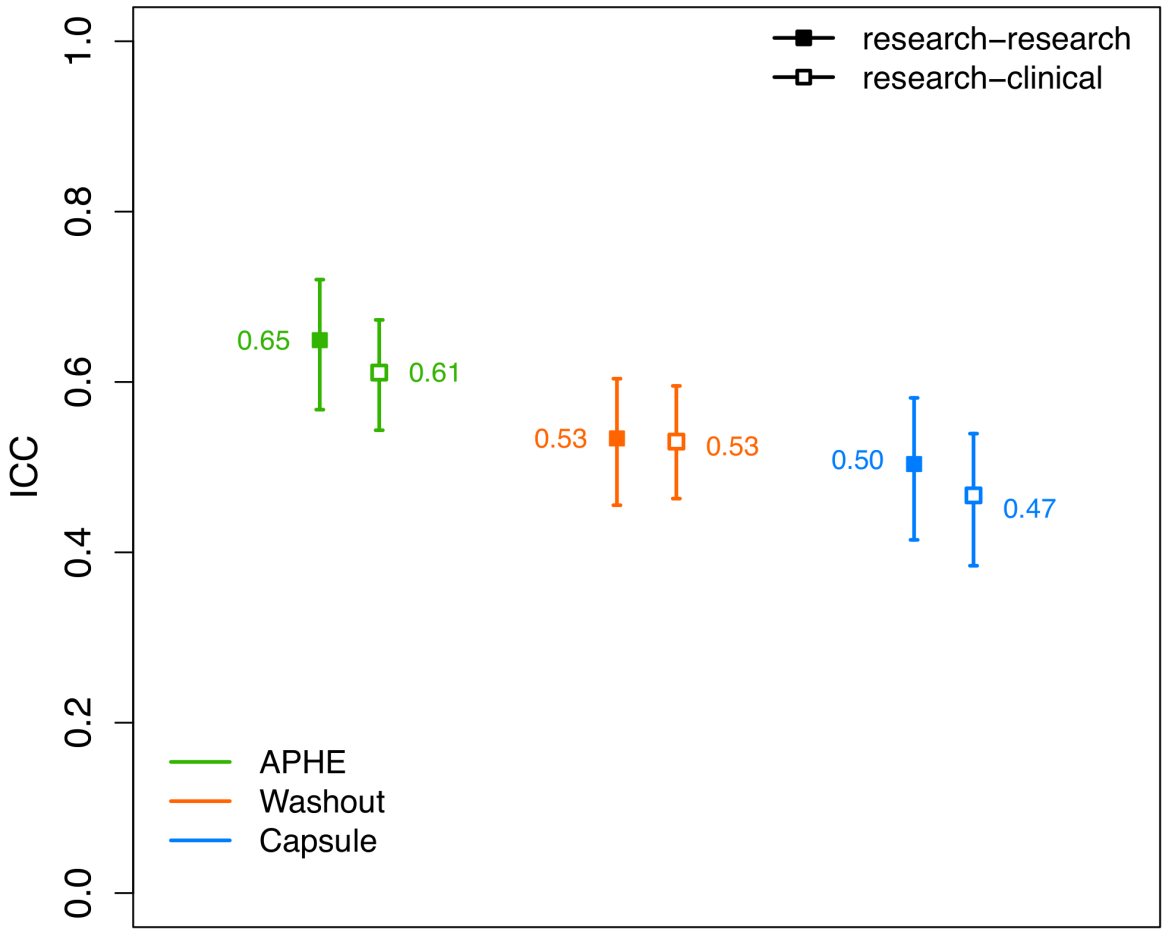


Figure 5. Reader agreement for LI-RADS version 2018 major features. Plot shows interclass correlation coefficients (ICC) for arterial phase hyperenhancement (green), washout (orange), and capsule (blue) for research reads only (filled squares) and between research and clinical reads (open squares) are shown. Tails represent 95% confidence intervals. No differences in ICCs between research-versus-research reads compared to researcher-versus-clinical reads were observed.

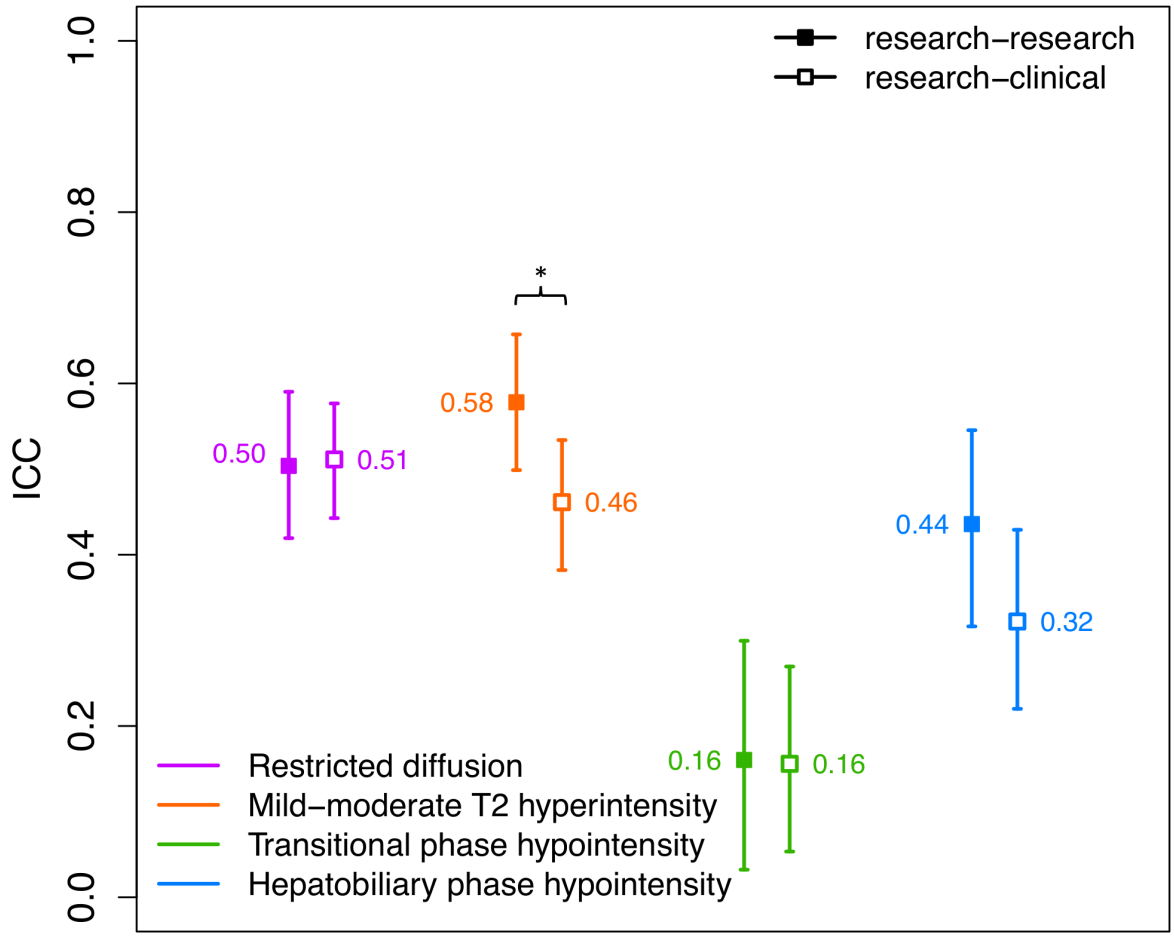


Figure 6. Reader agreement for LI-RADS version 2018 ancillary features. Plot shows interclass correlation coefficients (ICC) for ancillary features with sufficient frequency for analysis which included restricted diffusion (purple), mild-moderate T2 hyperintensity (orange), transitional phase hypointensity (green), and hepatobiliary phase hypointensity (blue). All are MRI-only features. Agreement among research reads only (filled squares) and between research and clinical reads (open squares) are shown. Tails represent 95% confidence intervals. *Represents a *P* value of < .05 by non-parametric bootstrap with per-case resampling. Research-versus-research agreement for mild-moderate T2 hyperintensity was better than research-versus-clinical agreement.

FIGURE LEGENDS

Figure 1: Schematic of the retrospective study design. De-identified examinations from the coordinating site and five other submitting sites were randomly assigned to two of 43 research readers for research reads. Features and category were extracted from the clinical report. This permitted the computation of inter-reader agreement between the research readers (R vs. R) and between the research and clinical readers (R vs. C). 20% of cases were also read twice by one of the research readers to permit the computation of intra-reader agreement (R' vs. R').

Figure 2. Reader agreement for modified four-category LI-RADS version 2018 scale based on imaging modality. Plot shows interclass correlation coefficients (ICC) for both modalities (blue), for MRI only (green), and for CT only (purple). Agreement among research reads only (filled squares) and between research and clinical reads (open squares) are shown. Tails represent 95% confidence intervals. *Represents a *P* value of < .05 by non-parametric bootstrap with per-case resampling. Research-versus-research agreement pooled over both modalities and for MRI only was better than research-versus-clinical agreement.

Figure 3. MR images showing examples where readers disagreed (A) and where readers agreed (B). The top row (A) shows images from an MR examination performed with gadoxetic acid from a 56-year-old male with cirrhosis secondary to hepatitis C. From left to right: pre-contrast (Pre), arterial phase (AP), portal venous phase (PVP), hepatobiliary phase (HBP). This 21 mm hepatobiliary phase hypointense observation (white arrow) was characterized on the clinical read as having non-rim arterial phase hyperenhancement and washout appearance and was categorized as LI-RADS category LR-5 (definitely hepatocellular carcinoma [HCC]). The first research reader characterized it as having a targetoid appearance and categorized it as LR-M (probably or definitely malignant, not specific for HCC). The second research reader characterized it as having no major features and paralleling the blood pool and categorized it as LR-2 (probably benign). It

was subsequently resected and found to be a well-differentiated HCC. The bottom row (B) shows images from an MR examination performed with an extracellular contrast agent from a 61-year-old female with cirrhosis secondary to hepatitis C. From left to right: pre-contrast (Pre), arterial phase (AP), portal venous phase (PVP), delayed phase (DP). This 31 mm observation (yellow arrow) in the caudate lobe was characterized on the clinical read as having arterial phase hyperenhancement, washout appearance, and capsule appearance was categorized as LI-RADS category LR-5 (definitely hepatocellular carcinoma). Both research readers also categorized this observation as LR-5. The patient passed away from intracranial hemorrhage a few months later.

Figure 4. Reader agreement for dichotomized classification of LI-RADS version 2018. Plot shows interclass correlation coefficients (ICC) for the following dichotomized categories: probably or definitely malignant vs not (blue), LR-5 (definitely hepatocellular carcinoma [HCC]) vs. not LR-5 (green), and LR-M (probably or definitely malignant, not specific for HCC) vs not LR-M (orange). Agreement among research reads only (filled squares) and between research and clinical reads (open squares) are shown. Tails represent 95% confidence intervals. *Represents a *P* value of < .05 by non-parametric bootstrap with per-case resampling. Research-research agreement for malignant categories was better than research-clinical agreement.

Figure 5. Reader agreement for LI-RADS version 2018 major features. Plot shows interclass correlation coefficients (ICC) for arterial phase hyperenhancement (green), washout (orange), and capsule (blue) for research reads only (filled squares) and between research and clinical reads (open squares) are shown. Tails represent 95% confidence intervals. No differences in ICCs between research-versus-research reads compared to researcher-versus-clinical reads were observed.

Figure 6. Reader agreement for LI-RADS version 2018 ancillary features. Plot shows interclass correlation coefficients (ICC) for ancillary features with sufficient frequency for

analysis which included restricted diffusion (purple), mild-moderate T2 hyperintensity (orange), transitional phase hypointensity (green), and hepatobiliary phase hypointensity (blue). All are MRI-only features. Agreement among research reads only (filled squares) and between research and clinical reads (open squares) are shown. Tails represent 95% confidence intervals. *Represents a *P* value of < .05 by non-parametric bootstrap with per-case resampling. Research-versus-research agreement for mild-moderate T2 hyperintensity was better than research-versus-clinical agreement.

TABLES

Table 1: Patient and examination characteristics.

Variables		**Counts or Mean (*n* = 484)**
Age (years)		62 ± 10
Sex	Male	328 (68%)
	Female	156 (32%)
Imaging modality	CT	93 (19%)
	MRI with ECA	174 (36%)
	MRI with HBA	217 (45%)
LI-RADS v2018 categories based on features extracted from the clinical report	LR-1	2 (0%)
	LR-2	35 (7%)
	LR-3	95 (20%)
	LR-4	153 (32%)
	LR-5	164 (34%)
	LR-M	27 (6%)
	LR-TIV	6 (1%)
	LR-NC	2 (0%)
LI-RADS major features from the clinical report	APHE	356 (74%)
	Washout	275 (57%)
	Enhancing capsule	117 (24%)
Submitting radiologist initials, institution, and country	C.W.H., University of California, San Diego, USA	118 (24%)
	V.C., Montefiore Medical Center, USA	211 (44%)
	S.L., University of California, Irvine, USA	52 (11%)
	J.L., University of Kentucky, USA	30 (6%)
	J.Y.C., Yonsei University, South Korea	65 (13%)
	D.A., Fundación Santa Fe de Bogotá, Colombia	8 (2%)

Notes.—Data represent number of examinations with percentages for categorical variables and mean ± SD for continuous variables. APHE = arterial phase hyperenhancement, ECA = extracellular agents, HBA = hepatobiliary agents, LI-RADS = Liver Imaging Reporting and Data System.

Table 2: Research Reader Characteristics based on questionnaire results.

Question	Response Options	Response Results
Country of primary affiliation	USA	33
	Canada	2
	Brazil	2
	China	1
	South Korea	1
	India	1
	Columbia	1
	Australia	1
	Italy	1
Abdominal imaging fellowship	Yes	41/43 (95%)
	No	2/43 (5%)
Self-identified as expert in liver imaging	Yes	39/43 (91%)
	No	4/43 (10%)
Post-training experience (years)		11 ± 6 (0–30)
Practice patterns		
Modalities used at institution	Almost all MRI	8/43 (19%)
	More MRI than CT	20/43 (47%)
	Approximately equal	10/43 (23%)
	More CT than MRI	5/43 (12%)
	Almost all CT	0
MRI contrast agents used at institution	Mostly extracellular agents	25/43 (58%)
	Approximately equal	5/43 (12%)
	Mostly gadoxetic acid	13/43 (30%)
Liver cancer tumor board member	Yes	36/43 (84%)
	No	7/43 (16%)
Research read characteristics		
Self-reported time spent per case (minutes)		12 ± 5 (2–30)

Notes.—For country of primary affiliation, the number of readers is shown. For all other categorical variables, the proportion and percentage of readers selecting each response option of the reader questionnaire is shown. Mean ± standard deviation is shown for continuous variables with ranges in parentheses.

Supplemental Tables

Supplemental Table 1: List of Readers and Characteristics

Reader Initials	Institutional Affiliation	Location	Years of Post-training Experience
V.C.	Montefiore Medical Center	New York, NY, USA	14
J.Y.C.	Yonsei University	Seoul, Republic of Korea	20
S.L.	University of California, Irvine	Orange, CA, USA	5
R.K.	University of California, Irvine	Orange, CA, USA	7
C.L.	University of Florida	Jacksonville, FL, USA	20
J.T.L.	University Of Kentucky	Lexington, KY, USA	10
J.W.O.	University Of Kentucky	Lexington, KY, USA	5
D.A.A.	Fundación Santa Fe de Bogotá University Hospital	Bogota, Colombia	15
M.M.	University of Michigan	Ann Arbor, MI, USA	10
M.S.D.	University of Michigan	Ann Arbor, MI, USA	10
W.M.	University of Michigan	Ann Arbor, MI, USA	4
A.R.	Mount Sinai West Medical Center	New York, NY, USA	2
S.C.L.	Icahn School of Medicine at Mount Sinai	New York, NY, USA	9
A.S.K.	Weill Cornell Medicine	New York, NY, USA	5
E.M.H.	Weill Cornell Medicine	New York, NY, USA	17
M.R.B.	Duke University	Durham, NC, USA	10
G.B.	University Hospital of Palermo	Palermo, Italy	16
M.L.D.	University of California, Los Angeles	Los Angeles, CA, USA	14

M.A.O.	University of California, San Francisco	San Francisco, CA, USA	6
A.T.	Université de Montréal	Montréal, Canada	13
M.C.	Université de Montréal	Montréal, Canada	3
A.F.	Oregon Health & Science University	Portland, OR, USA	13
E.A.C.	Centro de Diagnóstico por Imagem	João Pessoa, Brazil	5
M.T.C.	University of California, Davis	Sacramento, CA, USA	10
J.P.M.	University of California, Davis	Sacramento, CA, USA	30
B.K.	Radiology Limited	Tucson, AZ, USA	13
K.M.E.	MD Anderson	Houston, TX, USA	15
V.R.S.	MD Anderson	Houston, TX, USA	10
K.B.	MD Anderson	Houston, TX, USA	0
R.M.M.	Naval Medical Center San Diego	San Diego, CA, USA	9
N.H.	University of São Paulo	São Paulo, Brazil	7
S.B.	University of Kansas	Kansas City, KS, USA	5
R.A.	University of Kansas	Kansas City, KS, USA	9
K.G.	Sir. H. N. Reliance Foundation Hospital	Mumbai, India	12
C.K.	California Pacific Medical Center	San Francisco, CA, USA	10
A.K.	Massachusetts General Hospital	Boston, MA, USA	8
J.W.	The 3rd Affiliated Hospital, Sun Yat-sen University	Guangzhou, China	20
I.C.	Inland Imaging	Spokane, WA, USA	12
B.B.	Sutter Medical Group	Sacramento, CA, USA	20
M.G.	Austin Health	Melbourne, Australia	15
G.M.C.	University of California, San Diego	San Diego, CA, USA	10

D.T.M.	University of California, San Diego	San Diego, CA, USA	2
K.J.F.	University of California, San Diego	San Diego, CA, USA	10

Notes.—List of readers in the study with their initials and reported primary affiliation, location, and years of post-training experience at the time of the study

Supplemental Table 2: Distribution of examination modalities and contrast agents at each submitting site

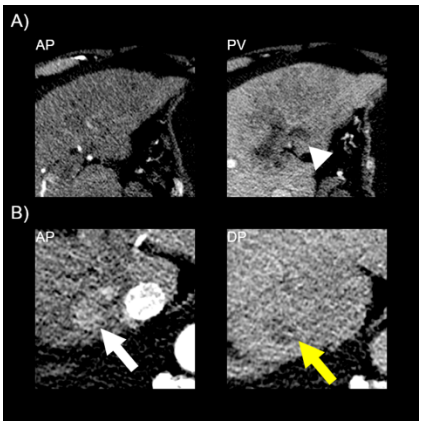
Site	CT	MRI with ECA	MRI with HBP
University of California, San Diego	19 (16%)	90 (76%)	9 (8%)
Montefiore Medical Center	48 (23%)	34 (16%)	129 (61%)
University of California, Irvine	20 (38%)	1 (2%)	31 (60%)
University of Kentucky	4 (13%)	26 (87%)	0
Yonsei University	0	19 (29%)	46 (71%)
Fundación Santa Fe de Bogotá	2 (25%)	4 (50%)	2 (25%)

Notes.—Number and row percentages of CT, MRI using extracellular agents (ECA), and MRI with hepatobiliary agents (HBP) from each submitting site

Supplemental Table 3

Site	Modality	Field Strength	Counts	Vendor(s)
University of California, San Diego	CT		19	GE (n=16), Toshiba (n=3)
	MRI	1.5T	39	GE (n=34), Siemens (n=2), Philips (n=2), Hitachi (n=1)
		3T	60	GE (n=60)
Montefiore Medical Center	CT		48	GE (n=46), Siemens (n=1), Toshiba (n=1)
	MRI	1.5T	99	Philips (n=98), Siemens (n=1)
		3T	64	Philips (n=62), GE (n=1), Siemens (n=1)
University of California, Irvine	CT		20	Philips (n=12), Siemens (n=8)
	MRI	1.5T	18	Siemens (n=18)
		3T	14	Philips (n=10), Siemens (n=4)
University of Kentucky	CT		4	Siemens (n=4)
	MRI	1.5T	25	Siemens (n=25)
		3T	1	GE (n=1)
Yonsei University	CT		0	-
	MRI	1.5T	0	-
		3T	65	Philips (n=44), Siemens (n=18), GE (n=3)
Fundación Santa Fe de Bogotá	CT		2	GE (n=2)
	MRI	1.5T	6	GE (n=6)
		3T	0	-

Notes.—Number of examinations that were performed using each vendor by modality and also by field strength for MRI from each submitting site.



Supplemental Figure 1: CT images showing examples where readers disagreed (A) and where readers agreed (B). The top row (A) shows arterial phase (AP) and portal venous (PV) images from a CT examination from a 82-year-old male with cirrhosis secondary to nonalcoholic steatohepatitis. This heterogeneous mass is associated with nonopacification of the left and main portal veins (arrowhead). This was categorized on the clinical read as LI-RADS category LR-M (probably or definitely malignant, not specific for hepatocellular carcinoma). The first research reader categorized it as LI-RADS category LR-5 (definitely HCC). The second research reader categorized it as LR-RADS category LR-TIV (tumor in vein). The patient transitioned to comfort care and passed away two months later. The bottom row (B) shows arterial phase (AP) and delayed phase (DP) images from a CT examination from a 69-year-old male with cirrhosis secondary to nonalcoholic steatohepatitis. This 18 mm observation in the hepatic dome was characterized on the clinical read as having arterial phase hyperenhancement (white arrow) and washout appearance (yellow arrow) and was categorized as LI-RADS category LR-5 (definitely HCC). Both research readers also categorized this observation as LI-RADS category LR-5. A follow-up examination four months later showed that the observation increased in size to 30 mm, the patient subsequently passed away from sepsis and hepatic encephalopathy eight months after that.

Supplemental Figure Legends

Supplemental Figure 1: CT images showing examples where readers disagreed (A) and where readers agreed (B). The top row (A) shows arterial phase (AP) and portal venous (PV) images from a CT examination from a 82-year-old male with cirrhosis secondary to nonalcoholic steatohepatitis. This heterogeneous mass is associated with nonopacification of the left and main portal veins (arrowhead). This was categorized on the clinical read as LI-RADS category LR-M (probably or definitely malignant, not specific for hepatocellular carcinoma). The first research reader categorized it as LI-RADS category LR-5 (definitely HCC). The second research reader categorized it as LR-RADS category LR-TIV (tumor in vein). The patient transitioned to comfort care and passed away two months later. The bottom row (B) shows arterial phase (AP) and delayed phase (DP) images from a CT examination from a 69-year-old male with cirrhosis secondary to nonalcoholic steatohepatitis. This 18 mm observation in the hepatic dome was characterized on the clinical read as having arterial phase hyperenhancement (white arrow) and washout appearance (yellow arrow) and was categorized as LI-RADS category LR-5 (definitely HCC). Both research readers also categorized this observation as LI-RADS category LR-5. A follow-up examination four months later showed that the observation increased in size to 30 mm, the patient subsequently passed away from sepsis and hepatic encephalopathy eight months after that.