Title: **Individual Participant Data Meta-Analysis of LR-5 in LI-RADS Version 2018 versus Revised LI-RADS for Hepatocellular Carcinoma Diagnosis**

Authors:

- Stacy M. Goins, BA. Duke University School of Medicine, Durham, NC, USA. Email: stacy.goins@duke.edu
- Hanyu Jiang, MD. Department of Radiology, West China Hospital, Sichuan University, Chengdu, China. Email: hanyu_jiang@foxmail.com
- Christian B. van der Pol MD, FRCPC. Juravinski Hospital and Cancer Centre, Hamilton Health Sciences, McMaster University, Hamilton, ON. Email: vanderpolc@hhsc.ca
- Jean-Paul Salameh BSc, MSc. Faculty of Health Sciences, Queen's University, Kingston, ON. Email: jeanp0409@gmail.com
- Eric Lam, M. Sc. Ottawa Hospital Research Institute, Ottawa, ON. Email: erlam@ohri.ca
- Robert G. Adamo, BSc, MA, MSc. Faculty of Medicine at The University of Ottawa, Ottawa, ON. Email: r.adamo@uottawa.ca
- Matthew DF McInnes MD PhD FRCPC, Departments of Radiology and Epidemiology uOttawa; The Ottawa Hospital Research Institute Clinical Epidemiology Program, Ottawa, ON. mmcinnes@toh.ca
- Andreu F. Costa, MD, MSc, FRCPC. Department of Diagnostic Radiology, Queen Elizabeth II Health Sciences Centre and Dalhousie University, Victoria General Building, 3rd floor, 1276 South Park Street, Halifax, Nova Scotia B3H 2Y9. andreufcosta@gmail.com
- An Tang, MD, MSc. Department of Radiology, Centre Hospitalier de l'Université de Montréal, Montréal, Québec, Canada. an.tang@umontreal.ca
- Ayman S. Alhasan, MD. Department of Radiology, College of Medicine, Taibah University, Medina, Saudi Arabia. Department of Radiology, King Faisal Specialist Hospital and Research Centre, Medina, Saudi Arabia. ahasan@taibahu.edu.sa
- Brian C. Allen, MD. Department of Radiology, Duke University Medical Center, Durham, NC. brian.allen@duke.edu
- Caecilia S. Reiner, MD. Institute of Diagnostic and Interventional Radiology, University Hospital Zurich, Zurich, Switzerland. caecilia.reiner@usz.ch

- Christopher Clarke, MBChB FRCR. Department of Radiology, Nottingham University Hospitals NHS Trust, Nottingham, United Kingdom. christopher.clarke@nuh.nhs.uk
- Milena Cerny, MD. Centre de Recherche du Centre Hospitalier de l'Université de Montréal, (CRCHUM), 900, rue Saint-Denis (Tour Viger), Montreal, QC H2X 0A9, Canada. milenacerny@gmail.com
- Jin Wang, MD. Department of Radiology, The Third Affiliated Hospital, Sun Yat-sen University (SYSU), Guangzhou, Guangdong, P.R. China. wangjin3@mail.sysu.edu.cn
- Sang Hyun Choi, MD PhD. Department of Radiology and Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea. edwardchoi83@gmail.com
- Tyler J. Fraum, MD. Mallinckrodt Institute of Radiology, Washington University School of Medicine, St Louis, MO. tylerjfraum@wustl.edu
- Daniel R. Ludwig, MD. Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, MO. ludwigd@wustl.edu
- Bin Song, MD. Department of Radiology, Sichuan University West China Hospital, Chengdu, Sichuan, China. songlab_radiology@163.com
- Ijin Joo, MD. Department of Radiology, Seoul National University College of Medicine, Jongno-gu, Seoul, Korea. hijijin@gmail.com
- Zhen Kang, MD. Department of Radiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Hubei, Wuhan, PR China. kangzh15@hust.edu.cn
- Andrea S. Kierans, MD. Weill Cornell Medical Center, New York City, NY, USA. ank9134@med.cornell.edu
- So Yeon Kim, MD. Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Songpa-Gu, Seoul, Republic of Korea. sykim.radiology@gmail.com
- Heejin Kwon, MD. Department of Radiology, Dong-A University Hospital, Dong-A University College of Medicine, Seogu, Busan, Republic of Korea. risual@dau.ac.kr
- Maxime Ronot, MD PhD. Department of Radiology, Hôpital Beaujon, APHP.Nord, Clichy & Université Paris Cité, CRI UMR 1149, Paris, France. maxime.ronot@aphp.fr
- Joanna Podgórska, MD. 2nd Radiology Department, Warsaw Medical University, Warsaw, Poland. jpodgo@gmail.com
- Grzegorz Rosiak, MD PhD. 2nd Radiology Department, Warsaw Medical University, Warsaw, Poland. grzegorzrosiak@yahoo.com

- Ji Soo Song, MD. Department of Radiology, Jeonbuk National University Medical School and Hospital, Deokjin-gu, Jeonju, Jeonbuk, South Korea. pichgo@gmail.com
- Mustafa R. Bashir, MD. Departments of Radiology and Medicine, Duke University Medical Center, Durham, NC, USA; Center for Advanced Magnetic Resonance Development, Duke University Medical Center, Durham, NC, USA; Department of Radiology, University of North Carolina, Chapel Hill, NC. Email: mustafa.bashir@duke.edu **(Corresponding Author)**

Original Research Location:

Duke University Health System, 40 Duke Medicine Circle, Durham, NC, 27710

Corresponding Author:

Mustafa R. Bashir, MD

40 Duke Medicine Circle, Durham, NC, 27710

Email: mustafa.bashir@duke.edu

Telephone: 919-684-2711

Funding:

Joan Sealy Trust for Cancer Research, Radiological Society of North America (RSNA) R&E Foundation Research Scholar Grant, CIHR Operating Research Grant, Fonds de recherche Quebec – Sante (FRQS #298509 to An Tang).

Manuscript Type:

Original Research, Meta-Analysis

Word Count: 3007

Data Sharing Statement:

Data generated by the authors or analyzed during the study are available in Supplementary Table 1.

Title: **Individual Participant Data Meta-Analysis of LR-5 in LI-RADS v2018 versus rLI-RADS for Hepatocellular Carcinoma Diagnosis**

Article Type: **Original Research, Meta-analysis**

Abbreviations:
HCC = Hepatocellular carcinoma
LI-RADS = Liver Imaging Reporting and Data System
rLI-RADS = Revised Liver Imaging Reporting and Data System
PPV = Positive predictive value
IPD = Individual-participant data
PRISMA-IPD = Preferred Reporting Items for a Systematic Review and Meta-Analysis of IPD
PRISMA-DTA = PRISMA Diagnostic Test Accuracy

Summary statement: **This individual-participant data meta-analysis confirmed that, compared to LI-RADS version 2018 category 5, the revised LI-RADS category 5 showed higher sensitivity and similar positive predictive value for diagnosing hepatocellular carcinoma.**

Key results:

1. In this meta-analysis of 24 studies including 3840 patients at high risk of hepatocellular carcinoma (HCC) and 4727 observations, the revised LI-RADS (rLI-RADS) category 5 (rLR-5) showed higher sensitivity (71% [95% CI: 61%, 79%] vs. 61% [95% CI: 46%, 75%]; $P < .001$) and similar positive predictive value (91% vs. 92%; $P = .55$) for HCC diagnosis compared to LI-RADS version 2018 category 5.
2. Category adjustments between LI-RADS v2018 and rLI-RADS were only assessed in categories 3, 4, and 5 (n=3619) as rLI-RADS only modifies these three categories, and the use of rLI-RADS increased ordinal category (LR-4 to rLR-5 or LR-3 to rLR-4) for 12% (432/3619) of observations.

ABSTRACT

Background:

A simplification of the Liver Imaging Reporting and Data Systems (LI-RADS) version 2018 (v2018), revised LI-RADS (rLI-RADS), has been proposed for imaging-based diagnosis of HCC. Single-site data suggest that rLI-RADS category 5 (rLR-5) improves sensitivity while maintaining positive predictive value (PPV) of the LI-RADS v2018 category 5 (LR-5), which indicates definite HCC.

Purpose:

To compare the diagnostic performance of v2018 and rLI-RADS in a multicenter dataset of patients at risk for HCC using individual-participant data (IPD) meta-analysis.

Methods:

Multiple databases were searched from December 2014 to January 2022 for studies evaluating the diagnostic performance of any version of LI-RADS on CT or MRI for diagnosing HCC. IPD meta-analysis methodology was applied to observations from the identified studies. QUADAS-2 was applied to determine study risk of bias. Observations were categorized according to major features and either v2018 or rLI-RADS assignments. Diagnostic accuracies of category 5 for each system were calculated using generalized linear mixed models and compared using the likelihood ratio test for sensitivity and the Wald test for PPV.

Results:

24 studies including 3840 patients and 4727 observations were analyzed. The median observation size was 19 mm (IQR: 11 mm – 30 mm). Compared to LR-5, rLR-5 showed higher sensitivity (70.6% [95% CI: 60.7%, 78.9%] vs. 61.3% [95% CI: 45.9%, 74.7%]; $P < .001$) with similar PPV (90.7% vs. 92.3%; $P = .55$). In low risk of bias studies (n = 4, 1031 observations), rLR-5 versus LR-5 also achieved a higher sensitivity (66.9% [95% CI: 58.2%, 74.5%] vs. 72.3% [95% CI: 63.9%, 80.1%]; $P = .02$) with similar PPV (88.7% vs 83.1%; $P = .47$).

Conclusion:

rLR-5 achieved a higher sensitivity for identifying HCC than LR-5 while maintaining a comparable PPV at ≥ 90%, matching the results presented in the original rLI-RADS study.

INTRODUCTION

Hepatocellular carcinoma (HCC) represents 75% -85% of all primary liver cancers and is the only cancer that can be definitively diagnosed on imaging alone for high-risk patients, those with cirrhosis or chronic hepatitis B virus (HBV) infection [1-3]. Numerous imaging-based diagnostic criteria for HCC have been established. The Liver Imaging Reporting and Data System (LI-RADS, current version 2018; v2018) is the predominant set of criteria used in the United States and has seen increasing adoption worldwide after recent endorsement by the American Association for the Study of Liver Disease [4-8].

Although LI-RADS v2018 (Figure 1A) is accepted as the standard for the imaging diagnosis of HCC in the United States, modifications have been proposed to improve the diagnostic performance of the LI-RADS 5 category (LR-5), which is considered definitely HCC on a scale from LR-1 to LR-5 [9-17]. One recently proposed modification of LI-RADS v2018 is the "revised LI-RADS" (rLI-RADS, Figure 1B) [10]. This set of criteria relies on gadoxetate disodium-enhanced MRI and is based on established combinations of major features in LI-RADS v2018. rLI-RADS was designed to improve the positive predictive value (PPV) for HCC and increase simplicity, with only 9 cells for categorization compared to 15 cells in v2018.

In the original study cohort, which included a test dataset of 55 patients and 195 observations, rLI-RADS category 5 showed increased sensitivity (75.5% vs. 60.9%; $P < .001$), decreased specificity (90.7% vs. 94.2%, $P = .008$), and similar PPV (92.5% vs. 94.1%; $P = .13$) compared with LI-RADS v2018 category 5 [10]. This dataset was derived from a single center and the majority of patients (52/55 [94%]) had HBV [10]. Thus, further evaluation of rLI-RADS in a multicenter setting is warranted to ensure the findings of the original study are generalizable.

The purpose of this study was to compare the performance of LI-RADS v2018 and rLI-RADS category 5 observations in a multicenter dataset of patients at risk for HCC. Sensitivity and PPV were the diagnostic performance measures of primary interest as sensitivity of category 5 for HCC represents the

proportion of HCCs that can be successfully diagnosed without additional testing and a PPV greater than or equal to 90% is considered sufficient for definite management [4].

METHODS

The study protocol was approved by the Research Ethics Board at the University of Ottawa and is Health Insurance Portability and Accountability Act compliant (Protocol details: [https://osf.io/duys4]). Best practices in diagnostic test accuracy systematic reviews were applied [18-20]. The Preferred Reporting Items for a Systematic Review and Meta-Analysis of Individual Patient Data (PRISMA-IPD) and PRISMA Diagnostic Test Accuracy (PRISMA-DTA) statements were used to inform reporting [21-24].

Study Design

The creation of this LI-RADS IPD is described in a previous publication *van der Pol et al.* [25]. Briefly, multiple databases were searched for studies published from January 2014 to December 2019 evaluating the diagnostic performance of LI-RADS for HCC [25]. The same methodology was used to update the search for studies published up to January 2022. Supplementary Material 1 and the Open Science Framework link above further detail the literature search, eligibility criteria, data collection and extraction, synthesis of results, and publication bias. Of note, the patients from the original rLI-RADS study were not included in the current work. Deidentified data were transferred to an encrypted Research Electronic Data Capture (REDCap) database at the University of Ottawa [26,27].

Risk of Bias and Applicability

The Quality Assessment of Diagnostic Accuracy Studies version 2 (QUADAS-2) tool for application to LI-RADS was used with signaling questions tailored as appropriate [28]. This previously tailored tool categorizes sources of bias into 4 domains – patient selection, index test, reference standard, and flow and timing. The flow and timing group included incomplete reporting of major features. Overall assessment of risk of bias was categorized as ‘low risk’ or ‘at risk’, and ‘at risk’ was determined if any of the 4 domains had ‘high’ risk of bias as per QUADAS-2 guidelines [28]. Assessment was performed in duplicate by two

independent authors, JPS and CvdP, both with five years of experience in conducting assessment of bias for studies in diagnostic test accuracy, and differences were resolved by a third author, MM, with 14 years of experience.

Conversion of LI-RADS categories across versions 2014, 2017, and 2018

Observations initially categorized according to LI-RADS v2014 and v2017 were updated to v2018 using a previously described conversion strategy [29].

Categorization of Observations

The proposed rLI-RADS differs from v2018 only in assigning observations to categories 3, 4, and 5 [4,10]. In this analysis, findings related to observations in all categories are reported to evaluate overall system performance and enable the findings of this study to be interpreted in the context of the existing literature [30-33]. Observations in categories 1, 2, 3, 4, M, and TIV were considered not definite HCC, as in prior studies [34-37]. When assessing differences in categorization between v2018 and rLI-RADS, however, only categories 3, 4, and 5 were considered as rLI-RADS does not affect any other category.

Observations were categorized in two ways: according to their reported major features based on either v2018 or rLI-RADS [4,10]. Of note, despite being underreported in the primarily literature, threshold growth was used as a major feature when available according to v2018 or rLI-RADS [38-40].

Diagnostic Performance Measures

The primary analysis focused on the diagnostic performances of category 5 (definitely HCC) in each system (i.e., LR-5 in v2018 and rLR-5 in rLI-RADS) for the diagnosis of HCC. The diagnostic measures included sensitivity, specificity, and PPV.

The same diagnostic performance measures were calculated in a sensitivity analysis using data only from studies determined to be at low risk of bias according to the QUADAS-2 tool. Another sensitivity analysis, limited to patients with HCC confirmed by histology, was also performed. Pre-specified sub-analyses were performed with patients stratified based on the presence or absence of a cirrhosis diagnosis and based on the types of contrast media (gadoxetate disodium vs. an extracellular agent) utilized for imaging (Supplementary Material 2).

Statistical Analysis

The performance of LR-5 and rLR-5 for the diagnosis of HCC was assessed using generalized linear mixed models (GLMMs). IPD were pooled across studies and modeled simultaneously with a one-stage meta-analysis approach to estimate pooled sensitivity and specificity, where the outcome was HCC status (positive or negative) and the exposure was whether a lesion was category 5 or other. A bivariate random effects model was fitted via maximum-likelihood estimation and used a binomial distribution with logit link [41]. Using the one stage meta-analytic approach, this models sensitivity and specificity simultaneously, accounting for the correlation between them and for precision of estimates within studies. For each analysis, this model provided estimates of pooled sensitivity and specificity. Study-level and patient-level effects were accounted for in the bivariate model through nested random effects [42].

PPVs for v2018 and rLI-RADS were calculated as estimates of the percentage of HCC observations in the LR-5 and rLR-5 categories, respectively. A separate univariate one-stage random effects model, clustering for study-level and patient-level nested random effects was used to calculate the pooled PPV of HCC [43]. Measures of heterogeneity included I^2, the percentage of the variability in effect estimates due to heterogeneity, and tau^2, the between-study variance, and were obtained through a maximum-likelihood estimator in the one-stage random effects model for PPV [41]. *P* values for sensitivity and specificity were calculated using a likelihood ratio test, and *P* values for PPV were calculated using the Wald test to compare

proportions of meta-analysis subgroups. All statistical analyses were performed using the glmer function in the 'lme4' package in R (version 4.0.0; R Foundation for Statistical Computing) [44]. The level of significance was set at $P < .05$.

.

RESULTS

Study Selection and Characteristics

The current study included 19 of 32 articles used in a previous IPD study [25]. Thirteen articles were excluded as either only contrast-enhanced ultrasound (CEUS) was performed or CT/MRI major features were not reported (Figure 2).

In the updated search for the current study, 407 articles were identified. Of those, 136 were excluded for the following reasons: participants were not adults at high risk for HCC; LI-RADS criteria for classification of observations was not used; the index test modality was not contrast-enhanced multiphase liver CT or MRI; and the reference standard was not met for diagnosis of HCC, other malignancy, or benign observation. After full-text review, 67 were excluded due to the same exclusion criteria. Ultimately, the authors of 204 studies were invited to collaborate. Thereafter, studies were excluded for the following reasons: no response from the author (n = 86), the author did not sign the data sharing agreement (n = 24), the author did not send the data (n = 54), the study did not follow LI-RADS guidelines for reference standards (n = 2), the study data were received after the inclusion deadline (n = 22), or the study did not report CT/MRI LI-RADS imaging characteristics (n = 1). Five additional studies from this updated search were included in this study. The final cohort from the original and updated search combined included a total of 4727 observations from 3840 patients in 24 studies (Figure 2).

Of the 24 studies, 2 reported features on CT, 17 evaluated MRI (of which 8 used gadoxetate exclusively and another 3 used gadoxetate and an extracellular agent), and 5 evaluated CT and MRI (4 used gadoxetate and an extracellular agent, and 1 used gadoxetate exclusively). Of the 4727 observations from these studies, 2195 were imaged with gadoxetate-enhanced MRI, 1620 were imaged with an extracellular agent on MRI, and 407 were imaged with CT. Of the 4727 observations, 397 (8.4%) from 7/24 (29.2%) studies reported threshold growth. Supplementary Table 1 further describes the characteristics of the included studies.

Risk of Bias and Applicability

Supplementary Table 2 summarizes the risk of bias and applicability for the studies included in this meta-analysis. Of the 24 studies, 21% (5/24) were limited to HCC observations, 21% (5/24) were limited to malignancies, and 13% (3/24) excluded benign observations.

In terms of risk of bias, 63% (15/24) of studies were unclear or had a high risk of bias for patient and observation selection, 33% (8/24) were unclear or had a high risk of bias under their index test (CT/MRI), 21% (5/24) were unclear or had a high risk of bias for the reference standard, and 63% (15/24) had a high risk of bias for study flow and timing. Additionally, there were inappropriate or unclear intervals between the index test and the reference standard in 21% (5/24) of the studies. Only 17% (4/24) of the studies, comprising 1031 observations, were considered to have a low risk of bias.

The percentage of variation across studies for analyses (I^2) ranged from 0.0% to 92.2%. The between-study variance (tau^2) ranged from 0.0 to 8.13.

Synthesis of Results

Observation Categorization

Based on v2018, 44% (2062/4727) of observations were categorized as LR-5, 12% (553/4727) as LR-4, and 21% (1004/4727) as LR-3 (Table 1). Based on rLI-RADS, 49% (2312/4727) of observations were categorized as rLR-5, 10% (485/4727) as rLR-4, and 17% (822/4727) as rLR-3. The median size of all observations was 19 mm (IQR: 11 mm – 30 mm).

Changes in categorization only increased from v2018 to rLI-RADS, which occurred for 12% (432/3619) of observations. Observations only changed by one category (i.e., LR-3 to rLR-4) (Figure 3). Of the observations that increased in category from LR-3 to rLR-4, 40% (72/182) were HCC. Of the observations that increased in category from LR-4 to rLR-5, 74% (186/250) were HCC.

Primary Analysis

Overall, rLI-RADS rLR-5 observations achieved a higher sensitivity than v2018 LR-5 observations for the diagnosis of HCC (70.6% vs. 61.3%, *P* < .001) but showed a reduced specificity (85.6% vs. 89.2%, *P* = .002) (Table 2). The PPV of rLR-5 for HCC (90.7% [95% CI: 80.1%, 95.9%]) was similar to the PPV of LR-5 for HCC (92.3% [95% CI: 82.3%, 96.9%], *P* = .55).

Sensitivity Analysis

In low risk of bias studies (4 studies with 1031 observations), v2018 categorized 40% (414/1031) of observations as LR-5, 13% (129/1031) as LR-4, and 30% (309/1031) as LR-3 (Table 3). rLI-RADS categorized 47% (480/1031) of observations as rLR-5, 10% (99/1031) as rLR-4, and 26% (273/1031) as rLR-3. The median observation size was 16 mm (IQR: 11 mm – 22 mm).

When the diagnostic performances of only low risk of bias studies were assessed, rLI-RADS rLR-5 achieved a higher sensitivity than v2018 LR-5 for HCC (72.3% vs. 66.9%, *P* = .02) but showed a reduced specificity (90.7% vs. 94.1%, *P* = .001) (Table 4). The PPV of rLR-5 for HCC (83.1% [95% CI: 79.5%, 86.2%]) was similar to the PPV of LR-5 for HCC (88.7% [95% CI: 85.2%, 91.4%], *P* = .47).

Additional Analyses

The sensitivity analyses for observations with a histological reference standard showed an increased sensitivity of rLR-5 compared with LR-5 for HCC (76.4% vs. 67.5%, *P* < .001) with similar PPVs (90.7% [95% CI: 67.3%, 97.9%] vs. 94.4% [95% CI: 74.1%, 99.0%], *P* = .50) (Supplementary Table 3).

For the sub-analyses with patients stratified based on the presence or absence of a cirrhosis diagnosis as well as based on contrast agent type (gadoxetate disodium vs. an extracellular agent), the finding of increased sensitivity and similar PPV of rLR-5 compared with LR-5 for HCC was again observed (Supplementary Table 4-7).

DISCUSSION

A recent study proposed modifications of the Liver Imaging Reporting and Data System (LI-RADS) version 2018 (v2018) major feature system for diagnosing hepatocellular carcinoma (HCC) [10]. Category 5 of this "revised LI-RADS" (rLI-RADS) showed higher sensitivity, lower specificity, and similar positive predictive value (PPV) for HCC diagnosis when compared with v2018 LR-5 [10]. The aim of our study was to perform an external validation comparing the performance of v2018 and rLI-RADS in an individual participant data (IPD) meta-analysis of a large multicenter dataset of patients at risk for HCC. Similar to the original study, we found that rLI-RADS category 5 (rLR-5) had a higher sensitivity and similar PPV for HCC when compared with v2018 category 5 (LR-5) across all sensitivity and subgroup analyses. Our primary analysis showed sensitivities of rLR-5 and LR-5 for HCC of 70.6% and 61.3% ($P < .001$), respectively, compared with 75.5% and 60.9% ($P < .001$) in the original description of rLI-RADS [10]. The specificities of rLR-5 and LR-5 for HCC in our study were 85.6% and 89.2% ($P = .002$), respectively, compared with 90.7% and 94.2% ($P = .008$) in the original description [10]. The PPVs of rLR-5 and LR-5 for HCC in our study were 90.7% and 92.3% ($P = .55$), respectively, compared with 92.5% and 94.1% ($P = .13$) in the original description [10].

The diagnostic measure of greatest interest in describing the performance of these criteria for diagnosing HCC is PPV [10,45]. Prior research has suggested that a PPV ≥ 90% is acceptable for identifying HCC and in fact is required to guarantee appropriate treatment, such as liver transplantation and locoregional therapies such as radioembolization, without additional diagnostic workup [10,45,46]. Given the consequences of treatment selection based on imaging diagnosis alone, a high and reliable PPV is critical, especially when therapies such as transplantation are an option. In the current study, the point estimates of the PPVs for category 5 lesions in identifying HCC for both rLI-RADS and v2018 were greater than 90%. However, when assessing only low risk of bias studies, the PPVs for category 5 lesions in identifying HCC for rLI-RADS and v2018 were 83.1% and 88.7%, respectively. In the low risk of bias cohort,

the 95% confidence interval for PPV for v2018 included the target 90% threshold, but the 95% confidence interval for rLI-RADS category 5 was below the 90% threshold. This suggests that the PPV requirement was not maintained by rLI-RADS in low risk of bias studies. Results of this sensitivity analysis should be interpreted with caution, however, as only 4 studies were at low risk of bias.

The specificity results of our study should also be interpreted with caution. This diagnostic measure was reported for completeness but is challenging to interpret as it represents a summary performance measure that considers all non-LR-5 categories to be "negative for HCC". In practice, however, those categories (i.e. LR-4, LR-3, LR-M) have different performance characteristics and clinical implications, and are not treated as a single unit. For example, according to LIRADS v2018, an LR-5 observation associated with TIV is considered 'definite HCC'. In a different context, this could be considered a true positive, but for the purposes of evaluating the diagnostic performance of the LR-5 category in isolation, we treated LR-5 TIV as "not definite HCC".

Our study had several limitations. First, the I^2 and tau^2 values, representing heterogeneity for PPV within and between studies, respectively, showed large amounts of statistical heterogeneity [47]. This suggests that the included studies are heterogeneous in design, particularly with regard to the characteristics of the study samples, increasing the variance of PPV between studies. This heterogeneity could have caused wider and overlapping confidence intervals in the pooled estimates of sensitivity and specificity. Second, since only 4 of the 24 included studies were determined to be at low risk of bias, there was a much smaller sample size for the sensitivity analysis performed on low risk-of-bias studies. The greatest potential sources of bias were those of the primary literature, including patient and observation selection, as well as study flow and timing [48-50]. This and other IPD meta-analyses have identified a need for higher-quality liver imaging research based on standard methods [25, 48-51]. Third, data collection was determined by the response of authors from the studies from which the data were included, which reduced the number of studies and thus the number of observations that could be analyzed. Finally,

the prevalence of LR-5 and rLR-5 lesions in this study is higher than seen in clinical practice, consistent with known selection biases in the primary literature.

In conclusion, we performed external validation in a multicenter cohort and found that rLI-RADS rLR-5 has higher sensitivity, lower specificity, and a similar PPV for HCC when compared with v2018 LR-5, confirming the findings from the original study. Our study also provides a platform for external performance assessment of other proposed modifications to the LI-RADS criteria. Future studies would benefit from using similar methodology to evaluate other proposed modifications of rLI-RADS to determine diagnostic performance and in turn help guide diagnosis of HCC in patients at high risk.

REFERENCES

1. Bray, F., et al., Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin, 2018. 68(6): p. 394-424.
2. Tsujita, Y., et al., Evaluation and Prediction of Treatment Response for Hepatocellular Carcinoma. Magn Reson Med Sci, 2023.
3. Park, J., et al., Imaging Diagnosis of HCC: Future directions with special emphasis on hepatobiliary MRI and contrast-enhanced ultrasound. Clin Mol Hepatol, 2021.
4. ACR, CT/MRI LI-RADS v2018 CORE. 2018.
5. Neri, E., et al., ESGAR consensus statement on liver MR imaging and clinical use of liver-specific contrast agents. Eur Radiol, 2016. 26(4): p. 921-31.
6. Hasegawa, K., et al., Clinical Practice Guidelines for Hepatocellular Carcinoma: The Japan Society of Hepatology 2021 version (5th JSH-HCC Guidelines). Hepatol Res, 2023.
7. Chernyak, V., et al., Liver Imaging Reporting and Data System (LI-RADS) Version 2018: Imaging of Hepatocellular Carcinoma in At-Risk Patients. Radiology, 2018. 289(3): p. 816-830.
8. Singal, A.G., et al., AASLD practice guidance on prevention, diagnosis, and treatment of hepatocellular carcinoma. Hepatology, 2023.
9. Jiang, H., et al., Modifying LI-RADS on Gadoxetate Disodium-Enhanced MRI: A Secondary Analysis of a Prospective Observational Study. J Magn Reson Imaging, 2022.
10. Jiang, H., et al., Data-Driven Modification of the LI-RADS Major Feature System on Gadoxetate Disodium-Enhanced MRI: Toward Better Sensitivity and Simplicity. J Magn Reson Imaging, 2022. 55(2): p. 493-506.
11. Rong, D., et al., Comparison of Gadobenate-Enhanced MRI and Gadoxetate-Enhanced MRI for Hepatocellular Carcinoma Detection Using LI-RADS Version 2018: A Prospective Intraindividual Randomized Study. AJR Am J Roentgenol, 2022. 218(4): p. 687-698.
12. Li, C.Q., et al., An assessment of liver lesions using a combination of CEUS LI-RADS and AFP. Abdom Radiol (NY), 2022. 47(4): p. 1311-1320.
13. Chen, J., et al., Increasing the sensitivity of LI-RADS v2018 for diagnosis of small (10-19 mm) HCC on extracellular contrast-enhanced MRI. Abdom Radiol (NY), 2021. 46(4): p. 1530-1542.
14. Becker, A.S., et al., Increased interreader agreement in diagnosis of hepatocellular carcinoma using an adapted LI-RADS algorithm. Eur J Radiol, 2017. 86: p. 33-40.
15. Xie, S., et al., Can modified LI-RADS increase the sensitivity of LI-RADS v2018 for the diagnosis of 10-19 mm hepatocellular carcinoma on gadoxetic acid-enhanced MRI? Abdom Radiol (NY), 2022. 47(2): p. 596-607.
16. Min, J.H., et al., A modified LI-RADS: targetoid tumors with enhancing capsule can be diagnosed as HCC instead of LR-M lesions. Eur Radiol, 2022. 32(2): p. 912-922.
17. Hwang, S.H., et al., Optimal lexicon of gadoxetic acid-enhanced magnetic resonance imaging for the diagnosis of hepatocellular carcinoma modified from LI-RADS. Abdom Radiol (NY), 2019. 44(9): p. 3078-3088.
18. McGrath, T.A., et al., Best practices for MRI systematic reviews and meta-analyses. J Magn Reson Imaging, 2019. 49(7): p. e51-e64.
19. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Cochrane Training. 2022.
20. McInnes, M.D. and P.M. Bossuyt, Pitfalls of Systematic Reviews and Meta-Analyses in Imaging Research. Radiology, 2015. 277(1): p. 13-21.
21. McGrath, T.A., et al., Recommendations for reporting of systematic reviews and meta-analyses of diagnostic test accuracy: a systematic review. Syst Rev, 2017. 6(1): p. 194.
22. Stewart, L.A., et al., Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. JAMA, 2015. 313(16): p. 1657-65.

23. McInnes, M.D.F., et al., Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA, 2018. 319(4): p. 388-396.
24. Salameh, J.P., et al., Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. BMJ, 2020. 370: p. m2632.
25. van der Pol, C.B., et al., CT/MRI and CEUS LI-RADS Major Features Association with Hepatocellular Carcinoma: Individual Patient Data Meta-Analysis. Radiology, 2022. 302(2): p. 326-335.
26. Harris, P.A., et al., Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform, 2009. 42(2): p. 377-81.
27. Harris, P.A., et al., The REDCap consortium: Building an international community of software platform partners. J Biomed Inform, 2019. 95: p. 103208.
28. Whiting, P.F., et al., QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med, 2011. 155(8): p. 529-36.
29. Goins, S.M., et al., Conversion Strategy for LI-RADS Category 5 Observations across Versions 2014, 2017, and 2018. Radiology, 2023: p. 222971.
30. Zhou, Y., et al., Risk Stratification and Distribution of Hepatocellular Carcinomas in CEUS and CT/MRI LI-RADS: A Meta-Analysis. Front Oncol, 2022. 12: p. 873913.
31. Li, L., et al., Clinical Application of Liver Imaging Reporting and Data System for Characterizing Liver Neoplasms: A Meta-Analysis. Diagnostics (Basel), 2021. 11(2).
32. Lee, S., et al., CT and MRI Liver Imaging Reporting and Data System Version 2018 for Hepatocellular Carcinoma: A Systematic Review With Meta-Analysis. J Am Coll Radiol, 2020. 17(10): p. 1199-1206.
33. Shin, J., et al., Diagnostic Performance of LI-RADS v2018 versus KLCA-NCC 2018 Criteria for Hepatocellular Carcinoma Using Magnetic Resonance Imaging with Hepatobiliary Agent: A Systematic Review and Meta-Analysis of Comparative Studies. Gut Liver, 2023. 17(3): p. 466-474.
34. Alhasan, A., et al., LI-RADS for CT diagnosis of hepatocellular carcinoma: performance of major and ancillary features. Abdom Radiol (NY), 2019. 44(2): p. 517-528.
35. Shao, S., et al., Diagnostic performance of LI-RADS version 2018 in differentiating hepatocellular carcinoma from other hepatic malignancies in patients with hepatitis B virus infection. Bosn J Basic Med Sci, 2020. 20(3): p. 401-410.
36. Zhang, N., et al., Does Training in LI-RADS Version 2018 Improve Readers' Agreement with the Expert Consensus and Inter-reader Agreement in MRI Interpretation? J Magn Reson Imaging, 2021. 54(6): p. 1922-1934.
37. Lee, S., et al., Percentages of Hepatocellular Carcinoma in LI-RADS Categories with CT and MRI: A Systematic Review and Meta-Analysis. Radiology, 2023. 307(1): p. e220646.
38. Park, J.H., et al., Should Threshold Growth Be Considered a Major Feature in the Diagnosis of Hepatocellular Carcinoma Using LI-RADS? Korean J Radiol, 2021. 22(10): p. 1628-1639.
39. Ko, A., et al., Comparison of the diagnostic performance of the 2017 and 2018 versions of LI-RADS for hepatocellular carcinoma on gadoxetic acid enhanced MRI. Clin Radiol, 2020. 75(4): p. 319 e1-319 e9.
40. Granata, V., et al., Critical analysis of the major and ancillary imaging features of LI-RADS on 127 proven HCCs evaluated with functional and morphological MRI: Lights and shadows. Oncotarget, 2017. 8(31): p. 51224-51237.
41. Riley R.D., et al., Meta-analysis of diagnostic test studies using individual patient data and aggregate data. Stat Med, 2008. 27: p. 6111-36.

42. Stephen, W., et al., Maximum Likelihood for Generalized Linear Models with Nested Random Effects via High-Order, Multivariate Laplace Approximation. Journal of Computational and Graphical Statistics, 2000. 9:1, p. 141-157.
43. Kosinski, A.S., A weighted generalized score statistic for comparison of predictive values of diagnostic tests. Stat Med, 2013. 32(6): p. 964-77.
44. Bates, D.M., M.; Bolker, B. M.; Walker, S. C., Fitting Linear Mixed-Effects Models using lme4. Journal of Statistical Software, 2015. 67: p. 1-48.
45. van der Pol, C.B., 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, 2019. 156(4): p. 976-986.
46. Sirlin, C.B.C., V.; Fowler, K. J.; Kono, Y, Chapter 8: LI-RADS Diagnostic Categories, in LI-RADS v2018 CT/MRI Manual, V.S. Chernyak, C.B., Editor.: ACR Radiology.
47. Deeks JJ, H.J., Altman DG Chapter 10: Analysing data and undertaking meta-analyses, in Cochrane Handbook for Systematic Reviews of Interventions (version 6.3), T.J. Higgins JPT, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, Editor. 2022: Cochrane, 2022.
48. Pan, J., et al., Do transition and hepatobiliary phase hypointensity improve LI-RADS categorization as an alternative washout: a systematic review and meta-analysis. Eur Radiol, 2022. 32(8): p. 5134-5143.
49. Liang, Y., et al., Diagnostic performance of LI-RADS for MRI and CT detection of HCC: A systematic review and diagnostic meta-analysis. Eur J Radiol, 2021. 134: p. 109404.
50. Shin, J., et al., LI-RADS Major Features on MRI for Diagnosing Hepatocellular Carcinoma: A Systematic Review and Meta-Analysis. J Magn Reson Imaging, 2021. 54(2): p. 518-525.
51. Shin, J., et al., Diagnostic Performance of the 2018 EASL vs. LI-RADS for Hepatocellular Carcinoma Using CT and MRI: A Systematic Review and Meta-Analysis of Comparative Studies. J Magn Reson Imaging, 2023.
52. Allen, B.C., et al., Comparison of Visualization Rates of LI-RADS Version 2014 Major Features With IV Gadobenate Dimeglumine or Gadoxetate Disodium in Patients at Risk for Hepatocellular Carcinoma. AJR Am J Roentgenol, 2018. 210(6): p. 1266-1272.
53. Cerny, M., et al., LI-RADS for MR Imaging Diagnosis of Hepatocellular Carcinoma: Performance of Major and Ancillary Features. Radiology, 2018. 288(1): p. 118-128.
54. Chen, J., et al., Liver Imaging Reporting and Data System Category 5: MRI Predictors of Microvascular Invasion and Recurrence After Hepatectomy for Hepatocellular Carcinoma. AJR Am J Roentgenol, 2019. 213(4): p. 821-830.
55. Kim, D.H., et al., Arterial subtraction images of gadoxetate-enhanced MRI improve diagnosis of early-stage hepatocellular carcinoma. J Hepatol, 2019. 71(3): p. 534-542.
56. Kim, D.H., et al., Combined Hepatocellular-Cholangiocarcinoma: Magnetic Resonance Imaging Features and Prognosis According to Risk Factors for Hepatocellular Carcinoma. J Magn Reson Imaging, 2021. 53(6): p. 1803-1812.
57. Lee, C.M., et al., Combined computed tomography and magnetic resonance imaging improves diagnosis of hepatocellular carcinoma </= 3.0 cm. Hepatol Int, 2021. 15(3): p. 676-684.
58. Choi, J.Y., et al., Comparison of gadoxetate disodium-enhanced MRI sequences for measuring hepatic observation size and its implication of LI-RADS classification. Abdom Radiol (NY), 2022. 47(3): p. 1024-1031.
59. Clarke, C.G.D., et al., Comparison of LI-RADS with other non-invasive liver MRI criteria and radiological opinion for diagnosing hepatocellular carcinoma in cirrhotic livers using gadoxetic acid with histopathological explant correlation. Clin Radiol, 2021. 76(5): p. 333-341.
60. Fraum, T.J., et al., Differentiation of Hepatocellular Carcinoma from Other Hepatic Malignancies in Patients at Risk: Diagnostic Performance of the Liver Imaging Reporting and Data System Version 2014. Radiology, 2018. 286(1): p. 158-172.

61. Fraum, T.J., et al., Assessment of primary liver carcinomas other than hepatocellular carcinoma (HCC) with LI-RADS v2018: comparison of the LI-RADS target population to patients without LI-RADS-defined HCC risk factors. Eur Radiol, 2020. 30(2): p. 996-1007.
62. Ludwig, D.R., et al., Hepatocellular carcinoma (HCC) versus non-HCC: accuracy and reliability of Liver Imaging Reporting and Data System v2018. Abdom Radiol (NY), 2019. 44(6): p. 2116-2132.
63. Kang, Z., et al., Digital subtract angiography and lipiodol deposits following embolization in cirrhotic nodules of LIRADS category >/=3. Eur J Radiol Open, 2019. 6: p. 106-112.
64. Lim, K., et al., Added value of enhanced CT on LR-3 and LR-4 observation of Gd-EOB-DTPA MRI for the diagnosis of HCC: are CT and MR washout features interchangeable? Br J Radiol, 2022. 95(1132): p. 20210738.
65. Jiang, H., et al., Man or machine? Prospective comparison of the version 2018 EASL, LI-RADS criteria and a radiomics model to diagnose hepatocellular carcinoma. Cancer Imaging, 2019. 19(1): p. 84.
66. Joo, I., Jeon, SK., et al., Combined hepatocellular cholangiocarcinoma: LI-RADS v2017 categorisation for differential diagnosis and prognostication on gadoxetic acid-enhanced MR imaging. European Radiology, 2019. **29**: p. 373-382.
67. Kierans, A.S., et al., Validation of Liver Imaging Reporting and Data System 2017 (LI-RADS) Criteria for Imaging Diagnosis of Hepatocellular Carcinoma. J Magn Reson Imaging, 2019. 49(7): p. e205-e215.
68. Kim, Y.Y., et al., Hepatocellular Carcinoma versus Other Hepatic Malignancy in Cirrhosis: Performance of LI-RADS Version 2018. Radiology, 2019. 291(1): p. 72-80.
69. Ronot, M., et al., Comparison of the accuracy of AASLD and LI-RADS criteria for the non-invasive diagnosis of HCC smaller than 3 cm. J Hepatol, 2018. 68(4): p. 715-723.
70. Rosiak, G., et al., Comparison of LI-RADS v.2017 and ESGAR Guidelines Imaging Criteria in HCC Diagnosis Using MRI with Hepatobiliary Contrast Agents. Biomed Res Int, 2018. 2018: p. 7465126.
71. Seo, N., et al., Optimal criteria for hepatocellular carcinoma diagnosis using CT in patients undergoing liver transplantation. Eur Radiol, 2019. 29(2): p. 1022-1031.
72. Song, J.S., et al., LI-RADS v2014 categorization of hepatocellular carcinoma: Intraindividual comparison between gadopentetate dimeglumine-enhanced MRI and gadoxetic acid-enhanced MRI. Eur Radiol, 2019. 29(1): p. 401-410.
73. Stocker, D., et al., Does quantitative assessment of arterial phase hyperenhancement and washout improve LI-RADS v2018-based classification of liver lesions? Eur Radiol, 2020. 30(5): p. 2922-2933.
74. Zhang, L., et al., The Role of Preoperative Dynamic Contrast-enhanced 3.0-T MR Imaging in Predicting Early Recurrence in Patients With Early-Stage Hepatocellular Carcinomas After Curative Resection. Front Oncol, 2019. 9: p. 1336.

Table 1. Summary of the number of observations, percentage of observations, and size of observations for v2018 LI-RADS and revised LI-RADS categories 3, 4, and 5 (n = 24 studies).

	LI-RADS v2018 (n=4727)			rLI-RADS (n=4727)		
	LR-5	**LR-4**	**LR-3**	**rLR-5**	**rLR-4**	**rLR-3**
# of observations	2062	553	1004	2312	485	822
% of observations	44%	12%	21%	49%	10%	17%
Size of observations (mm)	24 (16 – 34)	21 (10 – 27)	12 (9 – 15)	23 (15 – 32)	21 (12 – 27)	12 (9 – 16)

Notes.— The size of observation is reported as the median with the IQR in parentheses. LI-RADS = Liver Imaging Reporting and Data System; rLI-RADS = Revised Liver Imaging Reporting and Data System.

Table 2. Diagnostic performance of LI-RADS v2018 category 5 and revised LI-RADS category 5 for the diagnosis of hepatocellular carcinoma (n = 24 studies).

	LI-RADS v2018 LR-5 (n=2062)	rLI-RADS rLR-5 (n=2312)	P-value
Sensitivity	61.3% (45.9%, 74.7%)	70.6% (60.7%, 78.9%)	< .001[α]
Specificity	89.2% (80.9%, 94.2%)	85.6% (78.2%, 90.8%)	.002[α]
PPV	92.3% (82.3%, 96.9%)	90.7% (80.1%, 95.9%)	.55[T]
I^2	92.2% (89.6%, 94.1%)	91.8% (89.0%, 93.9%)	
Tau^2	5.1	4.4	

Notes.— Data in parentheses are 95% CIs. Diagnostic estimates were computed using generalized linear mixed models to include study-level and patient-level random effects. The I^2 statistic for heterogeneity and tau^2 were obtained through a maximum-likelihood estimator for positive predictive value (PPV). LI-RADS = Liver Imaging Reporting and Data System; rLI-RADS = Revised Liver Imaging Reporting and Data System.
[α]*P values* were calculated using the likelihood ratio test.
[T]The *P* value for comparing proportions of meta-analysis subgroups was calculated using the Wald test.

Table 3. Summary of the number of observations, percentage of observations, and size of observations for v2018 LI-RADS and revised LI-RADS categories 3, 4, and 5 in low risk of bias studies (n = 4).

	LI-RADS v2018 (n=1031)			rLI-RADS (n=1031)		
	LR-5	**LR-4**	**LR-3**	**rLR-5**	**rLR-4**	**rLR-3**
# of observations	414	129	309	480	99	273
% of observations	40%	13%	30%	47%	10%	26%
Size of observations (mm)	20 (15 – 26)	20 (14 – 25)	13 (9 – 16)	19 (15 – 25)	20 (14 – 24)	13 (8 – 16)

Notes.— The size of observation is reported as the median with the IQR in parentheses. LI-RADS = Liver Imaging Reporting and Data System; rLI-RADS = Revised Liver Imaging Reporting and Data System.

Table 4. Diagnostic performance of LI-RADS v2018 category 5 and revised LI-RADS category 5 for the diagnosis of hepatocellular carcinoma for low risk of bias studies (n = 4 studies).

	LI-RADS v2018 LR-5 (n=414)	rLI-RADS rLR-5 (n=479)	P-value
Sensitivity	66.9% (58.2%, 74.5%)	72.3% (63.9%, 80.1%)	.02[α]
Specificity	94.1% (87.7%, 97.3%)	90.7% (81.1%, 95.7%)	.001[α]
PPV	88.7% (85.2%, 91.4%)	83.1% (79.5%, 86.2%)	.47[T]
I^2	0.0% (0.0%, 84.7%)	7.9% (0.0%, 85.9%)	
Tau^2	0.0	0.0	

Notes.— Data in parentheses are 95% CIs. Diagnostic estimates were computed using generalized linear mixed models to include study-level and patient-level random effects. The I^2 statistic for heterogeneity and tau^2 were obtained through a maximum-likelihood estimator for positive predictive value (PPV). LI-RADS = Liver Imaging Reporting and Data System; rLI-RADS = Revised Liver Imaging Reporting and Data System.
[α]*P values* were calculated using the likelihood ratio test.
[T]The P value for comparing proportions of meta-analysis subgroups was calculated using the Wald test.

FIGURES

A.

Arterial phase hyperenhancement (APHE)		No APHE		Nonrim APHE		
Observation size (mm)		< 20	≥ 20	< 10	10-19	≥ 20
Count additional major features: • Enhancing "capsule" • Nonperipheral "washout" • Threshold growth	None	LR-3	LR-3	LR-3	LR-3	LR-4
	One	LR-3	LR-4	LR-4	LR-4 / LR-5	LR-5
	≥ Two	LR-4	LR-4	LR-4	LR-5	LR-5

B.

Arterial phase hyperenhancement (APHE)		No APHE		APHE (nonrim)		
Observation size (mm)		< 20	≥ 20	< 10	10-19	≥ 20
Count additional major features: • Enhancing "capsule" • Nonperipheral "washout" • Threshold growth	None	rLR-3	rLR-3	LR-3	LR-3	LR-4
	One	rLR-4	rLR-4	rLR-5	rLR-5	LR-5
	≥ Two	rLR-5	rLR-5	rLR-5	rLR-5	LR-5

C.

Arterial phase hyperenhancement (APHE)		No APHE	APHE (nonrim)	
Observation size (mm)		Any size	< 20	≥ 20
Count additional major features: • Enhancing "capsule" • Nonperipheral "washout" • Threshold growth	None	rLR-3	rLR-3	rLR-4
	One	rLR-4	rLR-5	LR-5
	≥ Two	rLR-5	rLR-5	LR-5

Figure 1. Diagnostic tables for (A) CT/MRI LI-RADS v2018, (B) the expanded version of rLI-RADS, and (C) rLI-RADS. The expanded version of rLI-RADS (B) demonstrates how the rLI-RADS criteria is simpler when compared to LI-RADS v2018.

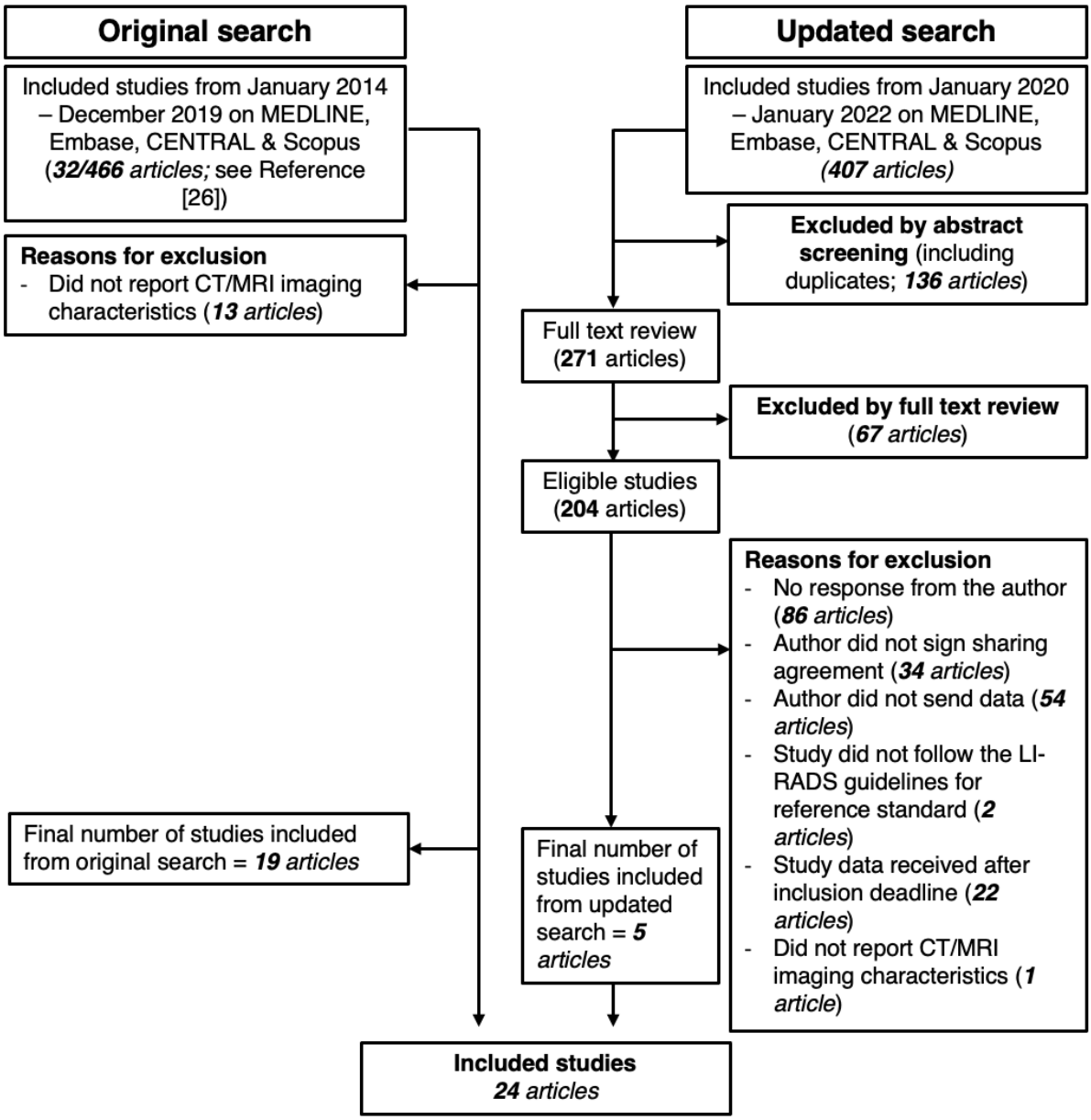

Figure 2. PRISMA flow diagram. Nineteen studies were included from the original search [26]. An additional five studies were included from the updated search, yielding a total of 24 studies included in this study.

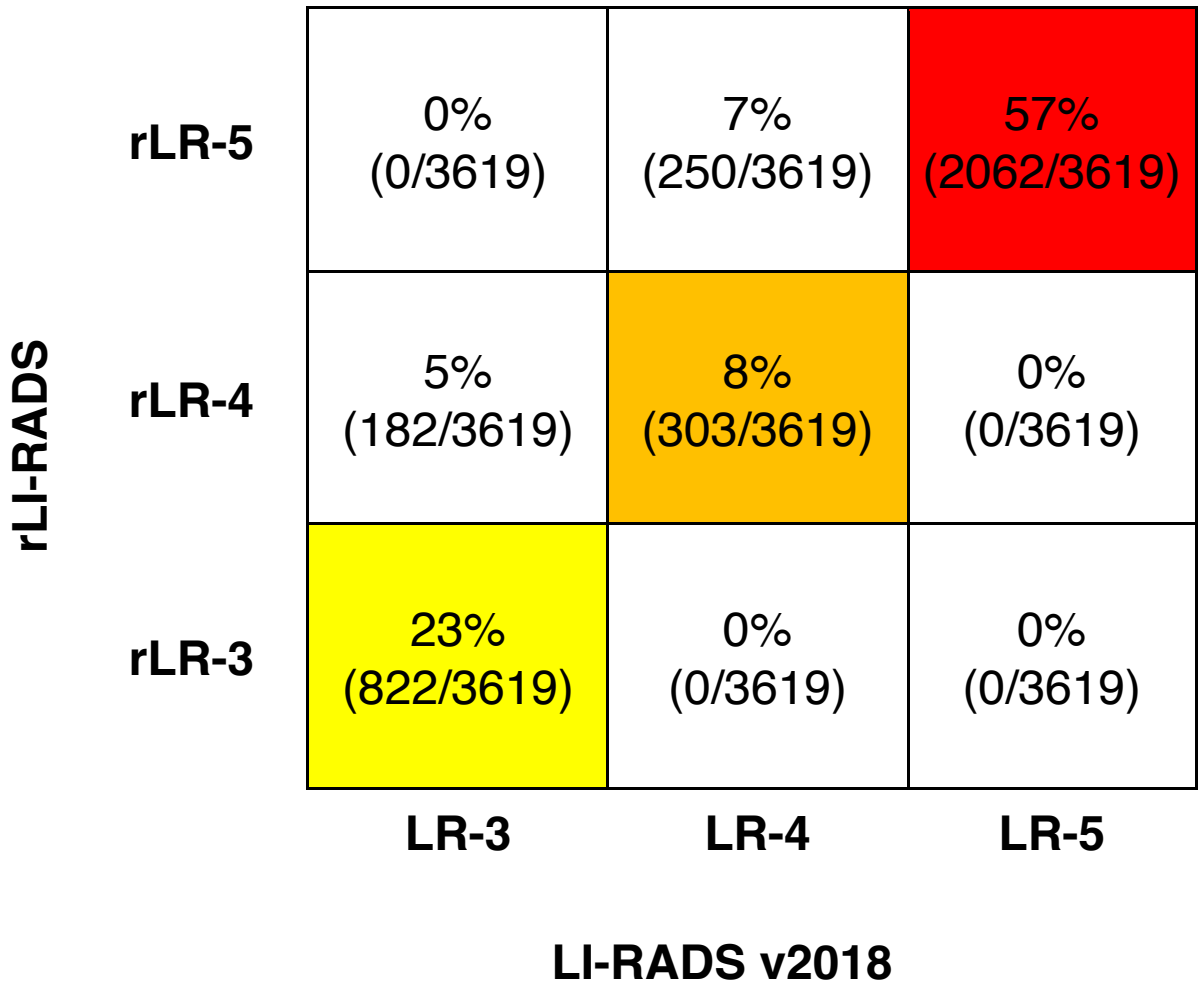

Figure 3. Migration of observation category between v2018 LI-RADS and rLI-RADS (n=24). Increases in observation category from v2018 to rLI-RADS occurred for 12% (432/3619) of observations in categories 3, 4, and 5 out of a total of 4727 observations in the entire cohort.

FIGURES

Figure 1. Diagnostic tables for (A) CT/MRI LI-RADS v2018 and (B) revised LI-RADS. The rLI-RADS criteria [10] are simpler than the LI-RADS v2018 criteria [4]. LI-RADS = Liver Imaging Reporting and Data System; rLI-RADS = Revised Liver Imaging Reporting and Data System.

Figure 2. PRISMA flow diagram. Nineteen studies were included from the original search of articles published from January 2014 to December 2019 described previously to create a LI-RADS individual patient database (IPD) [25]. An additional five studies were included from the updated search of articles published from January 2020 to January 2022 using the same methodology, yielding a total of 24 studies included in this study. LI-RADS = Liver Imaging Reporting and Data System; rLI-RADS = Revised Liver Imaging Reporting and Data System; Cochrane Central Register of Controlled Trials (CENTRAL).

Figure 3. Migration of observation category between v2018 LI-RADS and revised LI-RADS for 24 studies. Increases in observation category from v2018 to rLI-RADS occurred for 12% (432/3619) of observations in categories 3, 4, and 5 (shown in yellow, orange, and red, respectively) out of a total of 4727 observations in the entire dataset. Revised LI-RADS only differs from v2018 LI-RADS in categories 3, 4, and 5, so these were the only categories expected to change in category. LI-RADS = Liver Imaging Reporting and Data System; rLI-RADS = Revised Liver Imaging Reporting and Data System; Cochrane Central Register of Controlled Trials (CENTRAL).

SUPPLEMENTARY MATERIAL 1

Supplementary Methods:

Literature Search

An experienced hospital librarian assisted in the search of MEDLINE, Embase, Cochrane Central Register of Controlled Trials (CENTRAL), and Scopus databases for studies evaluating diagnostic performance of CT, MRI, or and contrast-enhanced ultrasound (CEUS) for diagnosis of HCC using LI-RADS (CT/MRI v2014/v2017/v2018 or CEUS LI-RADS v2016/v2017). There were no restrictions on language or publication type in the search. Dates included in the initial search were January 2014 to December 2019 based on publication date of LI-RADS 2014 [25]. The search strategy was updated to include additional eligible studies published from January 2020 to January 2022 (Figure 2). The corresponding authors of each study identified for inclusion were contacted.

Eligibility Criteria

Details of the search and process for inclusion in the Liver Imaging Reporting and Data System (LI-RADS) individual participant data (IPD) database are described in the study by *van der Pol et al.* [van 2022]. All imaging from patients at high risk of hepatocellular (HCC) including CT and MRI reporting the percentage of HCC for LI-RADS categories 3-5, were included. Patient are considered high risk for HCC if they have hepatic cirrhosis, chronic hepatitis B viral infection, current or prior HCC. Concordance with the LI-RADS technical imaging guidance for all CT, MRI, CEUS studies was evaluated. The use of CT/MRI LI-RADS v2014/v2017/v2018 or CEUS LI-RADS v2016/v2017 guidelines was required for all liver observations. Reference standard for diagnosis of HCC, other malignancy, or benign observations was assessed using a preferred composite reference standard [25].

Data Collection and Extraction

Authors were sent a single follow-up email if they did not respond to the initial invitation to collaborate. For those authors that agree to participate, a formal confidentiality agreement explaining secure data storage and authorized access by co-investigators was sent. This form included the data contribution form, data extraction sheet, data dictionary, and a list of frequently asked questions. Instructions to transfer data to an encrypted directory were also given to de-identify the data. Data sharing agreements were obtained per institution-specific policies. Patient data was not distributed otherwise. All collaborators were informed of progress as necessary.

Synthesis of Results

Data was collected in a master dataset in Research Electronic Data Capture (REDCap) with each observation being assigned a unique identifier [26]. IPD from primary study investigators and published reports were compared for each study. If data were shown to be unclear or inconsistent, the primary study investigators will be contacted for resolution. If there were multiple readers for a study, the data from one of the readers was randomly chosen. Notably, the data from which rLI-RADS was originally derived were part of the larger nodule database but were excluded from this analysis.

Publication Bias

Publication bias was not assessed as per contemporary guidance for DTA systematic reviews [23].

SUPPLEMENTARY MATERIAL 2

Supplementary Results:

A sensitivity analysis was performed in a cohort of patients whose HCC diagnosis was determined by pathology only (Supplementary Table 3). In this cohort, the sensitivity of rLR-5 for HCC was 76.4% and higher than that of LR-5 (67.5%, *P* < .001). The specificity of rLR-5 for HCC was 82.0% and lower than that of LR-5 (85.8%, *P* < .001). The PPVs of rLR-5 and LR-5 for HCC were 90.7% (95% CI: 67.3%, 97.9%) and 94.4% (95% CI: 74.1%, 99.0%; *P* = .50), respectively.

Sub-analyses were performed in a cohort of patients with cirrhosis versus a cohort of patients without cirrhosis for LR-5 and rLR-5 observations (Supplementary Table 4 and 5, respectively). In the cohort of patients with cirrhosis, the sensitivity of rLR-5 for HCC was 69.6% and higher than that of LR-5 (60.0%, *P* < .001). The specificity of rLR-5 for HCC was 85.1% and lower than that of LR-5 (88.4%, *P* = .006). The PPVs of rLR-5 and LR-5 for HCC were 91.4% (95% CI: 81.3%, 96.3%) and 92.0% (95% CI: 80.9%, 96.9%; *P* = .70), respectively. In the non-cirrhotic patient cohort, the sensitivity of rLR-5 for HCC was 82.9% and higher than that of LR-5 (75.7%, *P* = .003). The specificities of rLR-5 and LR-5 for HCC were 91.1% and 96.8%, respectively (*P* = .60). The PPVs of rLR-5 and LR-5 for HCC were 94.0% (95% CI: 77.1%, 98.6%) and 95.1% (95% CI: 77.9%, 99.1%; *P* = .72), respectively.

Sub-analyses were also performed in cohorts based on contrast media type (Supplementary Table 6 and 7, respectively). In the cohort imaged with gadoxetate disodium, the sensitivity of rLR-5 for HCC was 67.5% and higher than that of LR-5 (58.4%, *P* = .003). The specificities of rLR-5 and LR-5 for HCC were 87.0% and 89.1%, respectively (*P* = .67). The PPVs of rLR-5 and LR-5 for HCC were 92.0% (95% CI: 73.5%, 98.0%) and 92.6% (95% CI: 72.2%, 98.3%; *P* = .74), respectively. In the patient cohort imaged with an extracellular agent, the sensitivity of rLR-5 for HCC was 76.6% and higher than that of LR-5 (63.8%, *P* < .001). The specificity of rLR-5 for HCC was 83.6% and lower than that of LR-5 (89.3%, *P* = .002). The PPVs

of rLR-5 and LR-5 for HCC were 91.4% (95% CI: 79.0%, 96.8%) and 93.0% (95% CI: 79.2%, 97.9%; *P* = .56), respectively.

Supplementary Table 1. Characteristics of the included studies.

Study Details						Imaging Technique				Observation Data									
Author	Journal	Country	Design	Prevailing Risk Factor	No. of Cirrhotic Patients/No. of Patients	Modality	IV Contrast Agent Type	LI-RADS Version	No. of Readers	No. of Liver Observations/ No. of Patients	HCC	LR-1	LR-2	LR-3	LR-4	LR-5	LR-TIV	LR-M	Ref Standard
Alhassan 2019 [34]	Abdom Radiol (NY)	Canada	Retrospective Cohort	Cirrhosis >> HBV	55/59	CT	ECA	v2017	2	104/59	72	1	5	15	29	41	10	3	Pathology
Allen 2018 [52]	AJR Am J Roentgenol	USA	Retrospective case-control	Cirrhosis >> HBV	125/127	MRI	HPB	v2014	3	247/127	18	0	0	79	50	118	0	0	Pathology and CCRS
Cerny 2018 [53]	Radiology	Canada	Retrospective Cohort	Cirrhosis >> HBV	99/102	MRI	ECA	v2014	2	275/102	113	38	52	57	53	58	15	2	Pathology and CCRS
Chen 2019 [54]	AJR Am J Roentgenol	China	Retrospective Cohort	HBV > cirrhosis	57/149	MRI	ECA, HPB	v2018	2	149/149	149	0	0	0	0	149	0	0	Pathology
Kim DH 2019 [55]	J Hepatol	South Korea	Retrospective Cohort	HBV > cirrhosis	250/372	MRI	HPB	v2018	NR	372/258	273	0	0	18	154	180	4	16	Pathology and CCRS
Kim 2021 [56]	J Magn Reson Imaging	South Korea	Retrospective Cohort	Cirrhosis > HBV	65/113	MRI	Primovist	v2018	3	113/113	0	0	0	0	24	32	13	44	Pathology
Lee 2021 [57]	Hepatology International	South Korea	Retrospective Cohort	HBV >> cirrhosis	116/222	CT+MRI	Primovist	v2018	2	291/222	208	0	0	15	104	154	2	16	Pathology
Choi 2022 [58]	Abdom Radiol (NY)	South Korea	Retrospective Cohort	HBV >> cirrhosis	114/253	MRI	Primovist	v2018	2	279/253	247	0	0	18	46	190	6	19	Pathology and Clinical
Clarke 2021 [59]	Clinical Radiology	UK	Retrospective Cohort	Cirrhosis	47/47	MRI	HPB	v2018	2	105/47	70	0	0	29	38	36	2	0	Explant
Fraum 2018 [60]	Radiology	USA	Retrospective Cohort	Cirrhosis >> HBV	159/212	MRI	HPB	v2014	2	212/212	132	3	14	6	28	96	20	45	Pathology
Fraum 2020 [61, 62]*	European Radiology	USA	Retrospective Case-Control	Cirrhosis >> HBV	113/331	CT+MRI	ECA	v2018	2	331/331	81	0	0	2	12	63	24	230	Pathology
Kang 2019 [63]	Euro. Journal of Rad. Open	China	Retrospective Cohort	Cirrhosis >> HBV	19/19	MRI	ECA	v2014	2	19/19	15	0	0	4	2	11	1	1	Pathology and CCRS
Lim 2022 [64]	BR J Radiol	South Korea	Retrospective Cohort	Cirrhosis >> HBV	161/112	CT+MRI	CT: Ultravist MRI: Primovist	v2018	2	161/112	107	0	0	15	146	0	0	0	Pathology and CCRS
Jiang 2019 [65]	Cancer Imaging	China	Prospective Cohort	HBV >> cirrhosis	104/272	MRI	HPB	v2018	2	272/272	215	1	3	4	28	151	57	28	Pathology and CCRS
Joo 2017 [66]	European Radiology	South Korea	Retrospective Cohort	HBV >> cirrhosis	17/140	MRI	CT: ECA MRI:HPB	v2014	2	140/140	106	0	0	0	21	67	2	50	Pathology
Kierans 2018 [67]	J Magn Reson Imaging	USA	Retrospective Cohort	Cirrhosis > HBV	96/114	MRI	ECA, HPB	v2017	3	144/114	82	5	8	45	25	41	10	10	Pathology and CCRS
Kim YY 2019 [68]	Radiology	South Korea	Retrospective Cohort	Cirrhosis > HBV	220/220	MRI	HPB	v2018	2	220/220	165	0	0	5	10	70	135	0	Pathology
Ronot 2018 [69]	J Hepatol	France	Prospective Cohort	Cirrhosis >> HBV	422/422	CT+MRI	HPB	v2014	1	595/422	336	CT: 0 MRI: 0	CT: 61 MRI: 61	CT: 116 MRI: 133	CT: 195 MRI: 152	CT: 146 MRI: 207	CT: 0 MRI: 0	CT: 0 MRI: 0	Pathology and CCRS
Rosiak 2018 [70]	Biomed Res Int	Poland	Retrospective Cohort	Cirrhosis >> HBV	32/32	MRI	HPB	v2017	2	69/32	50	0	0	18	13	38	0	0	Pathology
Seo 2019 [71]	European Radiology	South Korea	Retrospective Cohort	HBV ~ cirrhosis	65/65	CT	ECA	v2014	2	R1: 67/50 R2: 102/65	R1: 42 R2: 54	R1:1 1 R2:1 6	R1: 1 R2: 18	R1:1 1 R2:1 4	R1:1 6 R2:2 1	R1:2 8 R2:3 1	NR	R1:0 R2:2	Pathology

Song 2019 [72]	European Radiology	South Korea	Retrospective Cohort	HBV ~ cirrhosis	108/154	MRI	ECA, HPB	v2014	2	154/154	154	0	0	2	64	88	0	0	Pathology and CCRS
Stocker 2020 [73]	European Radiology	SZ	Retrospective Cohort	Cirrhosis > HBV	53/60	MRI	ECA	v2018	4	71/60	28	18	11	15	6	21	0	0	Pathology and CCRS
van der Pol CB 2021 [25]	AJR Am J Roentgenol	Canada	Retrospective Cohort	HBV > cirrhosis	75/81	MRI	ECA, HPB	v2018	2	222/81	72	23	33	68	42	56	0	0	Pathology and CCRS
Zhang L 2019 [74]	Front. Oncol.	China	Retrospective Cohort	HBV ~ cirrhosis	82/82	MRI	ECA	v2018	2	82/82	82	0	0	7	7	68	0	0	Pathology

Notes. -- * Two authors and two studies contributed data for Fraum 2020, but there was data overlap and the Fraum 2020 study included all the data from both studies. Both are cited here for completeness. The > symbol indicates the first risk factor was more represented in the cohort than the second risk factor. The ~ symbol indicates both risk factors were represented approximately equally. The >> symbols indicate that the first risk factor was substantially more represented in the cohort than the second risk factor.

Abbreviations: LI-RADS, Liver Imaging Reporting and Data System; LR, Liver Imaging Reporting and Data System; HBV, Hepatitis B Virus; HCC, Hepatocellular Carcinoma; CT/MRI, Computed Tomography/Magnetic Resonance Imaging; USA, United States of America; UK, United Kingdom; ECA, Extracellular Contrast Agent; HPB, Hepatobiliary Contrast Agent; Ref, Reference; CCRS, Composite Reference Standard; NR, not recorded.

Supplementary Table 2. Assessment of risk of bias in the included studies.

Study (first author, year)	Patient and observation selection	Index CT	Index MRI	Reference standard	Flow and timing	Overall assessment	Applicability	Possible source of bias
Alhasan 2019 [34]	Low	Low	-	Low	High	At risk	Low concern	Interval between index test and reference standard
Allen 2018 [52]	High	-	Low	Unclear	High	At risk	Low concern	Case-control design, unclear interval between index test and reference standard. Limited to treated malignancies.
Cerny 2018 [53]	Low	-	Low	Low	High	At risk	Low concern	Interval between index test and reference standard
Chen 2019 [54]	High	-	Unclear	Low	High	At risk	Concerns	Limited to treated HCC categorized LR-5
Kim DH 2019 [55]	High	-	Unclear	Unclear	High	At risk	Low concern	Benign observations excluded, many methodology details unclear
Kim 2021 [56]	High	-	High	Low	Low	At risk	Low concern	Only path-proven malignancy was included, reviewers were aware of the final diagnosis
Lee 2021 [57]	High	-	Low	Low	High	At risk	Concerns	Only surgically resected patients included, likely biased towards more aggressive observations
Choi 2022 [58]	High	Low	Low	Low	Low	At risk	Low concern	Benign and probably benign observations were excluded

Clarke 2021 [59]	Low	-	Low	Low	High	At risk	Low concern	Benign lesions excluded from 2x2 table, prolonged interval from index test to ref standard in some cases
Fraum 2018 [60]	High	Low	Low	Low	Low	At risk	Low concern	Limited to malignancies
Fraum 2020 [61, 62]*	High	Low	Low	Low	Low	At risk	Low concern	Only malignant lesions included, some HCC randomly excluded, prolonged time from index test to reference standard
Kang 2019 [63]	Unclear	-	High	High	Low	At risk	Concerns	MRI technique and suboptimal reference standard
Lim 2022 [64]	Low	Low	Low	Low	High	At risk	Low concern	CT is at risk due to lack of delayed phase >3 min and reviewed with knowledge of MRI findings
Jiang 2019 [65]	Unclear	-	Low	High	High	At risk	Low concern	Almost exclusively resected observations
Joo 2017 [66]	High	High	High	Low	High	At risk	Low concern	Limited to HCC, readers aware of final diagnosis
Kierans 2019 [67]	Low	-	Low	Low	Low	Low risk	Low concern	None
Kim YY 2019 [68]	High	-	Low	Low	High	At risk	Low concern	Case-control design, limited to malignancies
Ronot 2018 [69]	Low	Low	Low	Low	Low	Low risk	Low concern	None

Rosiak 2018 [70]	High	-	Unclear	Low	High	At risk	Low concern	Limited to HCC, regenerative and dysplastic nodules
Seo 2019 [71]	Low	High	-	Low	High	At risk	Concerns	Limited to explanation, nonperipheral washout identified after data collection
Song 2019 [72]	High	-	High	Unclear	High	At risk	Low concern	Limited to HCC
Stocker 2020 [73]	Low	-	Low	Low	Low	Low risk	Low concern	None
van der Pol 2022 [25]	Low	-	Low	Low	Low	Low risk	Low concern	None
[74]	High	-	Low	Low	High	At risk	Low concern	Limited to resected HCC

Notes. -- Two authors and two studies contributed data for Fraum 2020, but there was data overlap and the Fraum 2020 study included all of the data from both studies. Both are cited here for completeness. Columns 2-6 are ranked 'low', 'high', or 'unclear' risk of bias. Column 7, for "Overall assessment", is ranked 'low risk' or 'at risk'. A study is considered 'at risk' for "Overall assessment" if any of the domains (columns 2-6) have 'high' risk of bias. Column 8, for "Applicability", is ranked 'low concern' or 'concerns'.

Supplementary Table 3. Diagnostic performance of LI-RADS v2018 category 5 and revised LI-RADS category 5 for the diagnosis of hepatocellular carcinoma in studies with only histology reference standard (n = 12 studies).

	LI-RADS v2018 LR-5 (n=1084/2396)	rLI-RADS rLR-5 (n=1222/2396)	P-value
Sensitivity	67.5% (42.2%, 85.5%)	76.4% (59.3%, 87.9%)	< .001[α]
Specificity	85.8% (61.2%, 95.9%)	82.0% (69.6%, 90.1%)	< .001[α]
PPV	94.4% (74.1%, 99.0%)	90.7% (67.3%, 97.9%)	.50[T]
I^2	71.2% (48.1%, 84.0%)	72.3% (50.4%, 84.5%)	
Tau^2	8.13	6.86	

Notes.— Data in parentheses are 95% CIs. Diagnostic estimates were computed using generalized linear mixed models to include study-level and patient-level random effects. The I^2 statistic for heterogeneity and tau^2 were obtained through a maximum-likelihood estimator for positive predictive value (PPV). LI-RADS = Liver Imaging Reporting and Data System; rLI-RADS = Revised Liver Imaging Reporting and Data System.
[α]*P values* were calculated using the likelihood ratio test.
[T]The P value for comparing proportions of meta-analysis subgroups was calculated using the Wald test.

Supplementary Table 4. Diagnostic performance of LI-RADS v2018 category 5 and revised LI-RADS category 5 for the diagnosis of hepatocellular in patients with cirrhosis.

	LI-RADS v2018 LR-5 (n=1426/3330)	rLI-RADS rLR-5 (n=1598/3330)	P-value
Sensitivity	60.0% (44.3%, 73.9%)	69.6% (59.8%, 77.8%)	< .001[α]
Specificity	88.4% (80.6%, 93.4%)	85.1% (77.4%, 90.5%)	.006[α]
PPV	92.0% (80.9%, 96.9%)	91.4% (81.3%, 96.3%)	.70[T]
I^2	91.2% (88.2%, 93.5%)	90.9% (87.6%, 93.3%)	
Tau^2	4.94	4.16	

Notes.— Data in parentheses are 95% CIs. Diagnostic estimates were computed using generalized linear mixed models to include study-level and patient-level random effects. The I^2 statistic for heterogeneity and tau^2 were obtained through a maximum-likelihood estimator for positive predictive value (PPV). LI-RADS = Liver Imaging Reporting and Data System; rLI-RADS = Revised Liver Imaging Reporting and Data System.
[α]*P values* were calculated using the likelihood ratio test.
[T]The P value for comparing proportions of meta-analysis subgroups was calculated using the Wald test.

Supplementary Table 5. Diagnostic performance of LI-RADS v2018 category 5 and revised LI-RADS category 5 for the diagnosis of hepatocellular in patients without cirrhosis.

	LI-RADS v2018 LR-5 (n=509/1058)	**rLI-RADS rLR-5 (n=547/1058)**	**P-value**
Sensitivity	75.7% (56.7%, 88.1%)	82.9% (69.9%, 91.0%)	.02[α]
Specificity	86.8% (67.3%, 95.4%)	91.1% (72.5%, 97.5%)	.60[α]
PPV	95.1% (77.9%, 99.1%)	94.0% (77.1%, 98.6%)	.72[T]
I^2	10.8% (0.0%, 48.0%)	22.9% (0.0%, 57.6%)	
Tau^2	7.57	6.67	

Notes.— Data in parentheses are 95% CIs. Diagnostic estimates were computed using generalized linear mixed models to include study-level and patient-level random effects. The I^2 statistic for heterogeneity and tau^2 were obtained through a maximum-likelihood estimator for positive predictive value (PPV). LI-RADS = Liver Imaging Reporting and Data System; rLI-RADS = Revised Liver Imaging Reporting and Data System.
[α]*P values* were calculated using the likelihood ratio test.
[T]The P value for comparing proportions of meta-analysis subgroups was calculated using the Wald test.

Supplementary Table 6. Diagnostic performance of LI-RADS v2018 category 5 and revised LI-RADS category 5 for the diagnosis of hepatocellular carcinoma in patients imaged with gadoxetate disodium.

	LI-RADS v2018 LR-5 (n=1049/2195)	rLI-RADS rLR-5 (n=1119/2195)	P-value
Sensitivity	58.4% (46.4%, 69.4%)	67.5% (60.4%, 74.0%)	.003[α]
Specificity	89.1% (74.9%, 95.8%)	87.0% (73.9%, 94.0%)	.67[α]
PPV	92.6% (72.2%, 98.3%)	92.0% (73.5%, 98.0%)	.74[T]
I^2	91.1% (87.0%, 93.9%)	91.4% (87.6%, 94.1%)	
Tau^2	7.46	6.83	

Notes.— Data in parentheses are 95% CIs. Diagnostic estimates were computed using generalized linear mixed models to include study-level and patient-level random effects. The I^2 statistic for heterogeneity and tau^2 were obtained through a maximum-likelihood estimator for positive predictive value (PPV). LI-RADS = Liver Imaging Reporting and Data System; rLI-RADS = Revised Liver Imaging Reporting and Data System.
[α]*P values* were calculated using the likelihood ratio test.
[T]The P value for comparing proportions of meta-analysis subgroups was calculated using the Wald test.

Supplementary Table 7. Diagnostic performance of LI-RADS v2018 category 5 and revised LI-RADS category 5 for the diagnosis of hepatocellular carcinoma in patients imaged with an extracellular agent.

	LI-RADS v2018 LR-5 (n=923/2169)	rLI-RADS rLR-5 (n=1051/2169)	P-value
Sensitivity	63.8% (37.5%, 83.9%)	76.6% (60.8%, 87.3%)	< .001[α]
Specificity	89.3% (76.7%, 95.5%)	83.6% (70.2%, 91.7%)	.002[α]
PPV	93.0% (79.2%, 97.9%)	91.4% (79.0%, 96.8%)	.56[T]
I^2	89.5% (84.2%, 93.1%)	88.1% (81.8%, 92.2%)	
Tau^2	4.38	3.22	

Notes.— Data in parentheses are 95% CIs. Diagnostic estimates were computed using generalized linear mixed models to include study-level and patient-level random effects. The I^2 statistic for heterogeneity and tau^2 were obtained through a maximum-likelihood estimator for positive predictive value (PPV). LI-RADS = Liver Imaging Reporting and Data System; rLI-RADS = Revised Liver Imaging Reporting and Data System.
[α]*P values* were calculated using the likelihood ratio test.
[T]The P value for comparing proportions of meta-analysis subgroups was calculated using the Wald test.