



Diagnosis of LI-RADS M lesions on gadoxetate-enhanced MRI: identifying cholangiocarcinoma-containing tumor with serum markers and imaging features

Hanyu Jiang^{1,2} · Bin Song¹ · Yun Qin¹ · Jie Chen¹ · Dong Xiao³ · Hong Il Ha² · Xijiao Liu¹ · Omobonike Oloruntoba-Sanders⁴ · Alaattin Erkanli³ · Andrew J. Muir⁴ · Mustafa R. Bashir^{2,4}

Received: 8 June 2020 / Revised: 14 September 2020 / Accepted: 6 November 2020 / Published online: 27 November 2020
© European Society of Radiology 2020

Abstract

Objectives The LI-RADS M (LR-M) category describes hepatic lesions probably or definitely malignant, but not specific for hepatocellular carcinoma in at-risk patients. Differentiation among LR-M entities, particularly detecting cholangiocarcinoma-containing tumors (M-CCs), is essential for treatment and prognosis. Thus, we aimed to develop diagnostic models on gadoxetate disodium-enhanced MRI comprising serum tumor markers and LI-RADS imaging features for M-CC.

Methods Consecutive at-risk patients with LR-M lesions exclusively (no co-existing LR-4 and/or LR-5 lesions) were retrieved retrospectively from a prospectively collected database spanning 3 years. Intrahepatic cholangiocarcinoma (ICC) and combined hepatocellular-cholangiocarcinoma (c-HCC-CCA) were classified together as M-CC. LI-RADS features determined by three independent radiologists and clinically relevant serum tumor markers were used to generate M-CC diagnostic models through logistic regression analysis against histology. Per-patient performance was evaluated using area under the receiver operating curve (AUC), sensitivity, and specificity.

Results Forty-five patients were included, 42.2% (19/45) with hepatocellular carcinoma, 33.3% (15/45) with ICC, 13.3% (6/45) with c-HCC-CCA, and 11.1% (5/45) with other hepatic lesions. Carbohydrate antigen (CA)19-9 > 38 U/mL, α -fetoprotein (AFP) > 4.8 ng/mL, and absence of the LI-RADS feature “blood products in mass” were significant predictors of M-CC. Combining three predictors demonstrated AUC of 0.862, sensitivity of 76%, and specificity of 88%. The risk of M-CC with all three criteria fulfilled was 98% (AUC, 0.690; sensitivity, 38%; specificity, 100%).

Conclusions In at-risk patients with LR-M lesions, integrating CA19-9, AFP, and the LI-RADS feature “blood products in mass” achieved high diagnostic performance for M-CC. When all three criteria were fulfilled, the specificity for M-CC was 100%.

Key Points

- In at-risk patients who had LR-M lesions exclusively (no concomitant LR-4/5 lesions), a model with carbohydrate antigen > 38 U/mL, α -fetoprotein > 4.8 ng/mL, and absence of the LI-RADS feature “blood products in mass” achieved high accuracy for diagnosing cholangiocarcinoma-containing tumors.
- In patients of whom all three criteria were fulfilled, the specificity for M-CC was 100%, which might reduce or eliminate the need for biopsy confirmation.

Keywords Diagnosis · Cholangiocarcinoma · Magnetic resonance imaging · Biomarkers · Tumor

Hanyu Jiang and Bin Song contributed equally to this work.

✉ Mustafa R. Bashir
mustafa.bashir@duke.edu

¹ Department of Radiology, West China Hospital, Sichuan University, Chengdu 610041, Sichuan, China

² Department of Radiology and Center for Advanced Magnetic Resonance Development, Duke University Medical Center, Box 3808, Durham, NC 27710, USA

³ Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC 27710, USA

⁴ Department of Medicine (Gastroenterology), Duke University Medical Center, Durham, NC 27710, USA

Abbreviations

AFP	α -Fetoprotein
CA	Carbohydrate antigen
c-HCC-CCA	Combined hepatocellular-cholangiocarcinoma
EOB-MRI	Gadoxetate disodium-enhanced magnetic resonance imaging
ICC	Intrahepatic cholangiocarcinoma
LI-RADS	Liver Imaging Reporting and Data System
M-CC	Cholangiocarcinoma-containing tumors

Introduction

In the latest version of the Liver Imaging Reporting and Data System (LI-RADS version 2018 [v2018]) [1], the LR-M category is used in at-risk patients to describe hepatic lesions which are probably or definitely malignant, but not specific for hepatocellular carcinoma (HCC). Accounting for approximately 36% and 30% of LR-M lesions [2–4], HCC and intrahepatic cholangiocarcinoma (ICC) are the two leading histologic tumor types within this category; less common etiologies include combined hepatocellular-cholangiocarcinoma (c-HCC-CCA, about 10% of the LR-M lesions), metastatic tumors, and others [2–5].

Differentiation between these primary liver malignancies is paramount in terms of treatment, prognosis, and survival. Specifically, the appropriateness of liver transplantation, optimal locoregional therapies, and chemotherapy choices differs substantially between cholangiocarcinoma-containing tumors (M-CC) and non-cholangiocarcinoma-containing tumors (non-M-CC) [6–10]. However, because of the overlapping risk factors, imaging characteristics, and clinical features between these entities within the LR-M category [6, 9–12], the diagnosis of LR-M lesions relies on histopathologic confirmation before treatment, which adds to costs of care and may give rise to undesired biopsy-related complications and treatment delays [1, 13].

To address this diagnostic dilemma, investigations have assessed the use of specific imaging features for distinguishing between different types of primary hepatic malignancies [14–21]. However, most of these studies are retrospective case-control in design without exclusive focus on LR-M lesions, instead including hepatic lesions representing a heterogeneous spectrum of diagnostic categories including LR-4, LR-5, LR-M, and LR-TIV. In addition, on review of the literature, we have found no study to date exploring the combined utility of imaging features and serum tumor markers in the assessment of patients with LR-M lesions, which could provide valuable complementary information [6, 9, 10, 22, 23]. In real-world settings, tumor markers are readily available and much less costly than imaging, but their interpretation in patients with LR-M lesions is not well established.

Therefore, the purpose of this study was to develop diagnostic models for M-CC incorporating predictive serum tumor markers and LI-RADS v2018 imaging features on gadoxetate disodium-enhanced magnetic resonance imaging (EOB-MRI) in at-risk patients with LR-M lesions.

Materials and methods

This single-center study was approved by our institutional review board, and the requirement for informed consent was waived because we retrospectively used data from a prospectively collected observational cohort (Clinical trial registration No. ChiCTR1900026668). All authors had access to the study data and reviewed and approved the final manuscript.

Patient sample

Between July 2015 and September 2018, consecutive adult at-risk patients with cirrhosis and/or chronic hepatitis B virus infection who underwent 3.0-T EOB-MRI for suspicious liver lesions detected by ultrasound and/or computed tomography were prospectively enrolled at our tertiary care hospital. According to the routine standard clinical practice of our center, EOB-MRI is usually performed in at-risk patients only when at least one suspicious liver lesion was detected by surveillance ultrasound and/or computed tomography performed for another indication. The diagnosis of cirrhosis was confirmed according to Chinese guidelines on the management of liver cirrhosis [24]. A study coordinator with 6 years of experience in liver imaging retrieved and de-identified the images, then reviewed and marked all measurable hepatic lesions with reference to the clinical reports. Exclusion criteria decided by the coordinator were as follows: any previous treatment for the hepatic lesion and MR imaging of insufficient quality for diagnosis.

All MR images were then independently reviewed by three abdominal radiologists (H.J., Y.Q., and X.L.) blinded to the final diagnosis with 5, 5, and 10 years of experience in liver imaging, respectively, using LI-RADS v2018. All patients in whom at least 2/3 radiologists identified LR-M lesions were included for further analyses. Patients were subsequently excluded at this level for co-existing LR-4 and/or LR-5 lesions because there was no way to separate the contributions to serum tumor marker levels from the LR-M lesions vs. the co-existing LR-4 and/or LR-5 lesions; no conclusive histopathologic result for the LR-M lesion within 1 month after initial MRI examination; no serum tumor marker results of interest within 14 days prior to histopathologic examinations. Of note, patients assigned as LR-TIV contiguous with LR-M lesions were included. Tumor markers of interest included α -fetoprotein (AFP), carcinoembryonic antigen (CEA), and carbohydrate antigen (CA) 19-9.

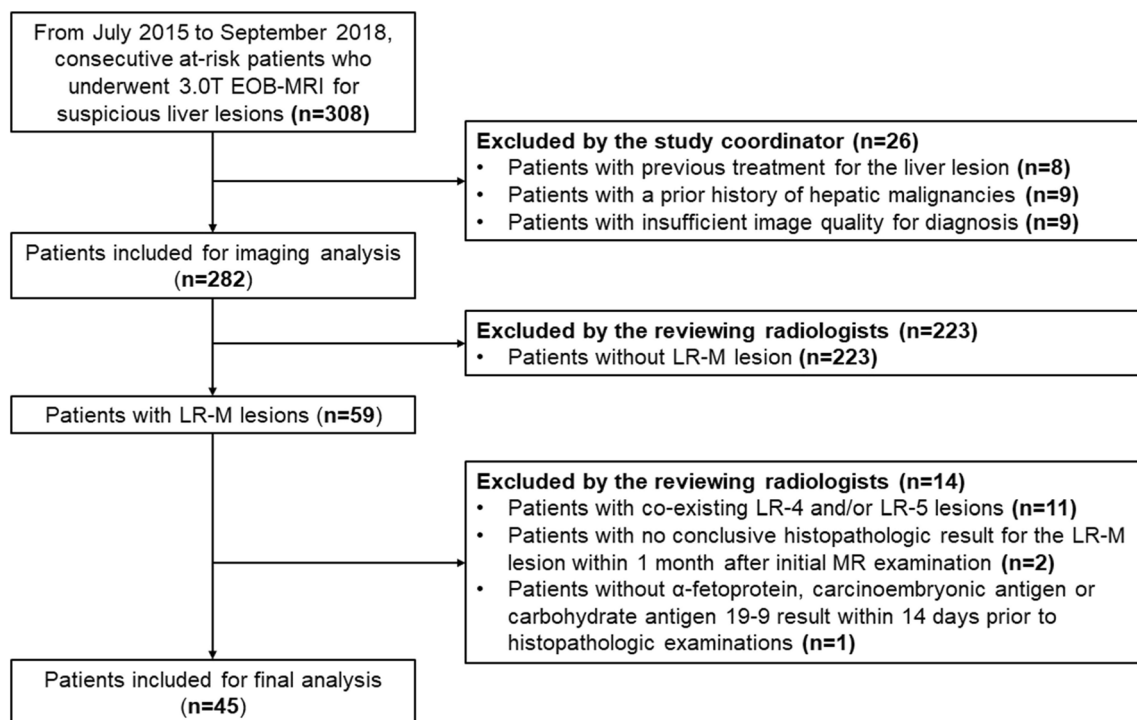


Fig. 1 Flow diagram of study sample selection

Of 308 patients who met the initial inclusion criteria, 263 were ultimately excluded (Fig. 1). The final study cohort included 45 patients (37 males; mean age 52.6 ± 9.6 years).

MRI protocols

All patients underwent EOB-MR examinations on a MAGNETOM Skyra 3.0-T MR scanner (Siemens Healthineers). MR sequences included the following: T2-weighted imaging; in- and opposed-phase T1-weighted sequence; diffusion-weighted imaging; and T1-weighted sequences before and after intravenous injection of EOB in the arterial phase, portal venous phase, transitional phase, and hepatobiliary phase. At the rate of 2 mL/s, 0.025 mmol/kg of EOB (Primovist; Bayer AG) was injected. Detailed MR protocols are presented in Supplementary Material 1.

Image analysis

For each included liver lesion, the same three reviewers as above independently evaluated the presence or absence of all LI-RADS v2018 major, ancillary, LR-M, and LR-TIV features (except those related to growth or US visibility, since those data were not available in the originating study database). For analyses requiring a single assessment, consensus was established by 2/3 majority based on the independent assessments.

All subsequent analyses were performed on a per-patient basis, and in patients with multiple LR-M lesions, the largest lesion was selected for feature-related analyses.

Reference standard

Histopathologic data, including hepatectomy and biopsy results, from routine report were used as the reference standard for determining tumor type. As a standard practice procedure, all specimens were reviewed by two independent pathologists (with 4 years and over 20 years of experience in liver oncology, respectively) who were aware of the clinical and imaging data, with disagreements resolved by discussion and consensus. All histopathologic diagnoses of the liver lesions were established according to the World Health Organization Classification [25].

For our analysis, as reported by and Park and colleagues [18], ICCs and c-HCC-CCAs were classified as M-CCs, while the remaining LR-M histologic types were classified as non-M-CCs.

Statistical analysis

Differences were compared with Student's *t* test or the Mann-Whitney *U* test for continuous variables, and with χ^2 test or Fisher's exact test for categorical variables, where applicable.

Feature selection and diagnostic models

Serum tumor markers and LI-RADS v2018 imaging features were selected according to stepwise algorithms to generate diagnostic models for M-CC on a per-patient basis as detailed below.

First, LI-RADS v2018 imaging features were selected based on the following: inter-rater agreement with Fleiss kappa ≥ 0.2 ; and prevalence of the feature in the dataset of at least 10% but no greater than 90%. Of note, features either too rare or too common were excluded considering their limited statistical utilities in differentiating between different LR-M entities to minimize potential model over-fitting. Along with four controlling factors (patient age, sex, underlying cirrhosis [present vs. absent], and lesion size), the remaining imaging features were entered into a multivariate logistic regression analysis using the stepwise method to construct a radiological model (Model R) for M-CC.

Second, for individual serum tumor markers, we used receiver operating curve analysis to evaluate their respective diagnostic performances and optimal threshold values for M-CC.

Finally, all three serum tumor markers and the imaging features meeting the inter-rater agreement and prevalence criteria described previously were entered into a multivariate logistic regression analysis to generate a fusion model (Model F) with adjustment for the same controlling factors.

Performances of the diagnostic models

Performances of the diagnostic models for M-CC were evaluated using area under the receiver operating curve (AUC), sensitivity, specificity, positive predictive value, negative predictive value, and accuracy. McNemar's and the DeLong tests were used to compare pairwise sensitivities, specificities, and AUCs, where applicable.

All statistical analyses were performed with statistical software (R version 3.3.1; R Foundation for Statistical Computing; or SPSS version 25. IBM). *P* value of less than 0.05 was considered statistically significant.

Results

Patients

Key characteristics of the included patients and liver lesions are summarized in Table 1 and Supplementary Material 2. In total, 88.9% (40/45) of the included patients were chronic hepatitis B virus carriers and 26.7% (12/45) had established cirrhosis. Respectively, 46.7% (21/45) and 53.3% (24/45) patients were diagnosed with M-CC and non-M-CC.

Among all included patients, 60.0% (27/45) had single LR-M lesions, while 40.0% (18/45) had multiple LR-M lesions (number of lesions, 2–8 per patient). Of 113 total observed LR-M lesions (median size, 17 mm; interquartile range, 10–53 mm), the 45 largest lesions were included in subsequent analyses (median size, 59 mm; interquartile range, 42–83 mm). No significant difference in size of the dominant lesions between patients with M-CC and non-M-CC ($p = 0.55$) was observed.

M-CCs were associated with significantly higher levels of CA19-9 (39.3 U/mL vs. 15.9 U/mL, $p = 0.043$) compared with non-M-CCs. However, there were no significant differences in AFP (9.2 ng/mL vs. 4.3 ng/mL, $p = 0.553$) or CEA levels (2.3 ng/mL vs. 2.2 ng/mL, $p = 0.776$, Table 1). Among M-CCs, significantly higher levels of AFP were observed among patients with c-HCC-CCA than those with ICC (504.3 ng/mL vs. 4.9 ng/mL, $p < 0.001$, Supplementary Material 2).

Diagnostic models and performance

Radiological model

According to the selection algorithm detailed in Fig. 2 and Table 2, LI-RADS v2018 imaging features were removed based on low inter-rater agreement ($n = 22$) and prevalence less than 10% or greater than 90% ($n = 3$). Three features were left after these two steps, but “mosaic architecture” and “tumor in vein” dropped out of the model at multivariate logistic regression analysis. Therefore, only the feature “blood products in mass” was eventually included in Model R, which demonstrated substantial inter-rater agreement (Fleiss kappa = 0.71), was present in 36% (16/45) of lesions, and had an odds ratio of 0.053 ($p = 0.006$) for M-CC. Model R (no blood products in mass) had an AUC of 0.655, sensitivity of 81% (17/21), specificity of 50% (12/24), and accuracy of 64% (29/45).

Fusion model

Ultimately, Model F was constructed with three predictors: no blood products in mass; AFP > 4.8 ng/mL; and CA19-9 > 38 U/mL (Fig. 2). A total of 18% (8/45), 42% (19/45), 96% (43/45), and 4% (2/45) patients had all three, two or more, one or more, and none above predictors present, respectively. The corresponding risks of M-CC were 98%, 86%, 49%, and 1% in these categories, respectively (Table 3).

The overall AUC, sensitivity, specificity, and accuracy of Model F for M-CC were 0.862, 76% (16/21), 88% (21/24), and 82% (37/45), respectively. Detailed performance characteristics for all models are summarized in Table 4.

Comparisons between model performances

Model F demonstrated significantly superior AUC (0.862 vs. 0.655, $p = 0.003$), specificity (88% vs. 50%, $p = 0.002$), and similar sensitivity (76% vs. 81%, $p = 0.706$) compared with **Model R** (Table 4).

The performances of Model F variants using specific numbers of present predictors are summarized in Table 4. The variant using two or more predictors present performed significantly better than the variants using three predictors present ($p = 0.047$) or one or more predictors present ($p < 0.001$) based on AUC.

Table 1 Clinical-pathologic characteristics of patients and liver lesions

Characteristic	Full cohort	M-CC	Non-M-CC	<i>p</i> value*
Patients	45	21 (47)	24 (53)	0.534
Age (years)	52.6 ± 9.6	54.3 ± 9.2	51.1 ± 10.2	0.277
Sex				
Male	37 (82)	18 (86)	19 (80)	0.574
Female	8 (18)	3 (14)	5 (21)	
Cirrhosis	12 (27)	5 (24)	7 (29)	0.686
Etiology of chronic liver disease				
Hepatitis B virus infection	40 (89)	18 (86)	22 (92)	0.555
Hepatitis C virus infection	2 (4)	2 (10)	0	
Hepatitis B and C virus co-infection	1 (2)	0	1 (4)	
Alcoholic liver disease	1 (2)	1 (5)	0	
Other causes	1 (2)	0	1 (4)	
Reference standard				
Resection	44 (98)	20 (95)	24 (100)	0.285
Biopsy	1 (2)	1 (5)	0	
Per-patient final diagnosis				
HCC	19 (42)	0	19 (79)	
ICC	15 (33)	15 (71)	0	
c-HCC-CCA	6 (13)	6 (29)	0	
Metastasis	2 (4)	0	2 (8)	
Other primary liver malignancies	2 (4)	0	2 (8)	
Benign lesions	1 (2)	0	1 (4)	
Serum tumor markers				
AFP (ng/mL)	5.5 (2.8–184.2)	9.2 (3.3–175.0)	4.3 (2.7–999.7)	0.553
CA19-9 (U/mL)	23.7 (11.3–42.6)	39.3 (17.0–385.2)	15.9 (10.7–30.9)	0.043
CEA (ng/mL)	2.3 (1.3–4.5)	2.3 (1.2–4.5)	2.2 (1.4–5.5)	0.776
Time intervals				
Between MRI and histopathologic diagnosis(d)	2 (1–4)	2 (1–4)	2 (1–5)	0.180
Between serum test and histopathologic diagnosis(d)	2 (3–5)	2 (3–5)	3 (2–6)	0.635
LR-M lesions				
No. of lesions	113	60 (53)	53 (47)	0.352
Solitary	27	11 (41)	16 (59)	0.142
Multiple	86	49 (57)	37 (43)	
Size (mm)†	59 (42–83)	59 (50–76)	56 (36–89)	0.554

Unless stated otherwise, data are the number of patients or lesions. Data are presented as count (percentage), mean ± standard deviation, and median (interquartile range), respectively

*Comparisons are made using Student's *t* test or Mann-Whitney *U* test for continuous variables, and with the χ^2 test or Fisher's exact test for categorical variables, where applicable

†Data described size of the largest lesion per patient

HCC hepatocellular carcinoma, ICC intrahepatic cholangiocarcinoma, cHCC-CCA combined HCC-cholangiocarcinoma, M-CC cholangiocarcinoma-containing tumors, AFP α -fetoprotein, CEA carcinoembryonic antigen, CA19-9 carbohydrate antigen 19-9, MRI magnetic resonance imaging

Representative cases of non-M-CC and M-CC are shown in Figs. 3 and 4, respectively.

Discussion

Despite remaining a challenging area in liver imaging, the prediction of histological tumor type for LR-M lesions may be enhanced by the assessment of serum tumor markers. By analyzing 45 at-risk patients with LR-M lesions, we found that combining CA19-9 value > 38 U/mL and AFP > 4.8 ng/mL with the LI-RADS v2018 feature no “blood products in mass” on EOB-MRI was the best available model for detecting M-CC lesions (sensitivity of 76%, specificity of 88%). When using three predictors present

as the cutoff value, the model provided excellent specificity (100%) for M-CC lesions, though with reduced sensitivity (38%).

Although substantial overlap was found between the imaging features of M-CC and non-M-CC tumors, “blood products in mass” was identified as a significant predictor favoring non-M-CC in our study. Defined as an ancillary feature “favoring HCC in particular” by LI-RADS v2018 [1], presence of “blood products in mass” is suggestive of rapid growth of hypervascular tumors which are prone to hemorrhage [26]. This feature has been shown useful in differentiating between HCC and non-HCC malignancies in at-risk patients in other studies [19, 27]. In contrast, most M-CCs are characterized by prominent desmoplastic and hypovascularized tumor stroma [7, 9], therefore rarely presenting with internal hemorrhage [26].

Table 2 Results and distributions of LI-RADS v2018 imaging features of the largest lesion per patient using consensus data

Imaging features	M-CC	Non-M-CC	Fleiss kappa value	Prevalence	Multivariate logistic analysis	
					OR	<i>p</i> value
Major features						
Nonrim APHE	2/21 (10)	4/24 (17)	0.057	6/45 (13.3)
Nonperipheral “washout”	1/21 (5)	4/24 (17)	0.098	5/45 (11.1)
Enhancing “capsule”	12/21 (57)	14/24 (58)	− 0.110	26/46 (57.8)
Ancillary features						
Restricted diffusion	21/21 (100)	24/24 (100)	...	45/45 (100.0)
Mild–moderate T2 hyperintensity	21/21 (100)	24/24 (100)	...	45/45 (100.0)
Corona enhancement	21/21 (100)	19/24 (79)	− 0.062	40/46 (88.9)
Fat sparing in solid mass	1/21 (5)	1/24 (4)	1.000	2/45 (4.4)
Iron sparing in solid mass	0/21 (0)	0/24 (0)	...	0/45 (0.0)
TP hypointensity	20/21 (95)	22/24 (92)	0.406	42/45 (93.3)
HBP hypointensity	20/21 (95)	24/24 (100)	0.492	44/45 (97.8)
Nonenhancing “capsule”	0/21 (0)	0/24 (0)	− 0.015	0/45 (0.0)
Nodule-in-nodule	2/21 (10)	5/24 (21)	0.074	7/45 (15.5)
Mosaic architecture	9/23 (43)	12/24 (50)	0.329	21/45 (46.7)	0.596	...
Blood products in mass	4/21 (19)	12/24 (50)	0.706	16/45 (35.6)	0.053	0.006
Fat in mass, more than adjacent liver	0/21 (0)	2/24 (8)	0.085	2/45 (4.4)
Parallels blood pool	0/21 (0)	0/24 (0)	...	0/45 (0.0)
Undistorted vessels	0/21 (0)	0/24 (0)	...	0/45 (0.0)
Iron in mass, more than liver	0/21 (0)	0/24 (0)	− 0.015	0/45 (0.0)
Marked T2 hyperintensity	0/21 (0)	0/24 (0)	...	0/45 (0.0)
HBP isointensity	0/21 (0)	0/24 (0)	...	0/45 (0.0)
LR-M features						
Rim APHE	18/21 (86)	19/24 (79)	0.056	37/45 (82.2)
Peripheral “washout”	6/21 (29)	0/24 (0)	0.037	6/45 (13.3)
Delayed central enhancement	9/21 (43)	10/24 (42)	0.097	19/45 (42.2)
Targetoid restriction	7/21 (33)	5/24 (21)	0.026	12/45 (26.7)
Targetoid TP or HBP appearance	6/21 (29)	5/24 (21)	0.019	11/45 (24.4)
Infiltrative appearance	11/21 (52)	12/24 (50)	0.022	23/45 (51.1)
Necrosis or severe ischemia	10/21 (48)	10/24 (42)	0.140	20/45 (44.4)
LR-TIV feature						
Tumor in vein	11/21 (52)	9/24 (38)	0.663	20/45 (44.4)	0.406	...

Data are number of the largest lesions per patients included. Data are presented as count (percentage)

M-CC cholangiocarcinoma-containing tumors, *OR* odds ratio, *APHE* arterial phase hyperenhancement, *TP* transitional phase, *HBP* hepatobiliary phase

Previous works have reported promising performances of LR-M imaging features, in particular targetoid appearance, for distinguishing between HCC and other liver malignancies in general at-risk patients, though not necessarily those with LR-M lesions specifically [14–21]. Other imaging features, including capsule appearance, intralesional fat, peritumoral bile duct dilatation, surface retraction, septum, and T2 hyperintense foci, have been variably found useful in some publications but not others [15, 18, 20, 21]. However, without being framed specifically in the context of LR-M lesions, these

findings can be difficult to apply. Our results demonstrated very limited incremental value of any of the individual LR-M imaging features in discriminating between different tumor types within the LR-M category. It was noteworthy that, to avoid substantial model over-fitting, a large proportion of features (79%) were excluded due to poor to slight inter-rater agreement. Nevertheless, agreement levels on LI-RADS imaging features were generally higher in previous studies [14–21]. This discrepancy could have been a result of the limited sample size of the current study. Additionally, rather

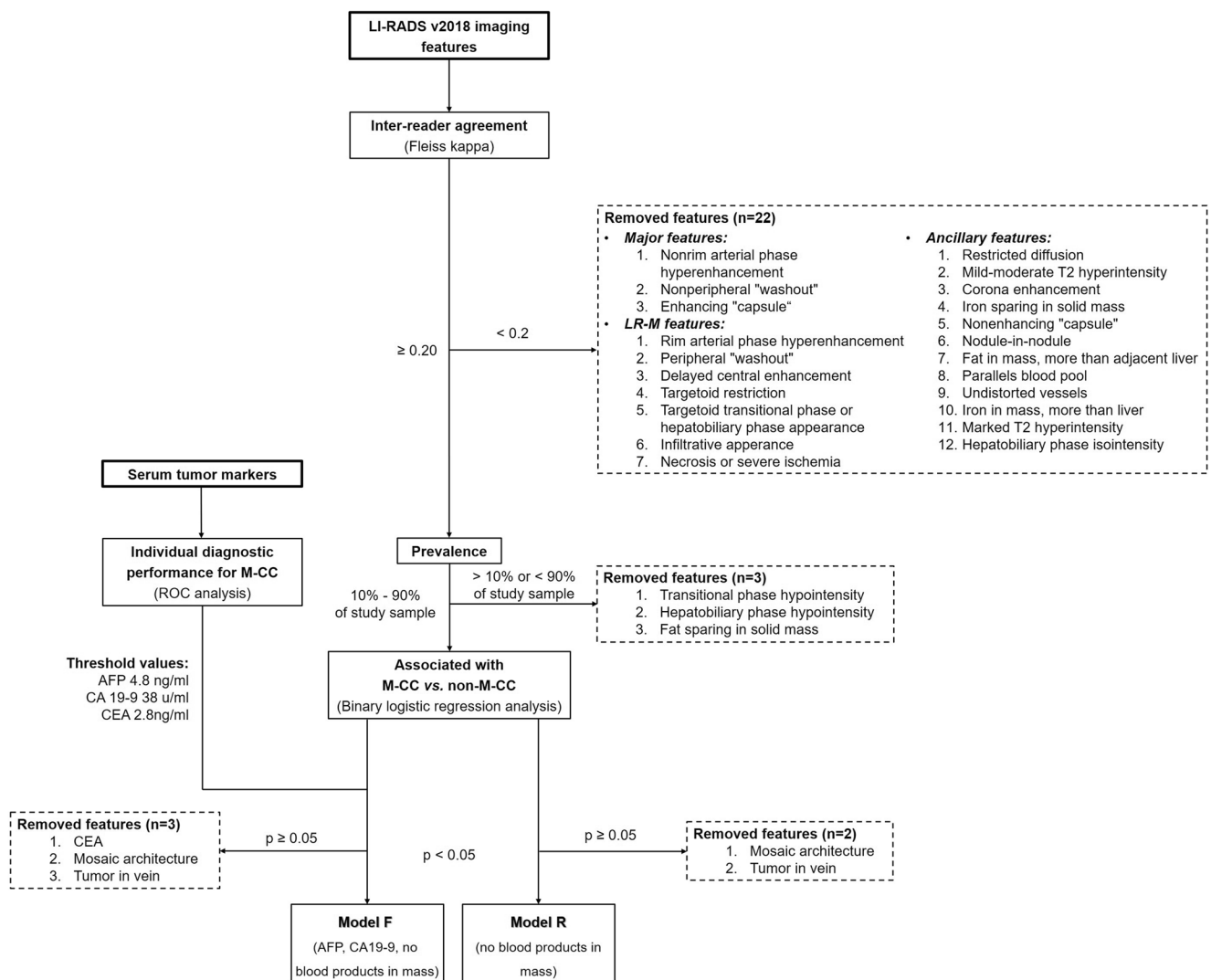


Fig. 2 Workflow of generating the diagnostic models for LR-M

Table 3 Risk of M-CC according to Model F

No blood products in mass	AFP > 4.8 ng/mL	CA19-9 > 38 U/mL	M-CC risk (by feature combination)	No. of patients (by feature combination)	Number of features present	M-CC risk (by number of features)	No. of patients (by number of features)
+	+	+	98%	8 (18)	Three	98%	8 (18)
+	+	-	86%	4 (9)	≥ Two	86%	19 (42)
+	-	+	75%	4 (9)			
-	+	+	69%	3 (7)	≥ One	49%	43 (96)
+	-	-	22%	13 (29)			
-	+	-	18%	10 (22)			
-	-	+	10%	1 (2)	None	1%	2 (4)
-	-	-	1%	2 (4)			

Data are presented as count (percentage)

AFP α-fetoprotein, CEA carcinoembryonic antigen, CA19-9 carbohydrate antigen 19-9, M-CC cholangiocarcinoma-containing tumors

Table 4 Per-patient performances of the diagnostic models for M-CC using consensus data

Diagnostic model	AUC	95% CI	No. of patients				Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)
			TP	FN	FP	TN					
Overall performances of Model R and Model F											
Model R	0.655	0.493–0.816	17	4	12	12	81	50	59	75	64
Model F	0.862	0.751–0.973	16	5	3	21	76	88	84	81	82
Specific performances of Model F using different cutoffs											
3 predictors present	0.690	0.530–0.851	8	13	0	24	38	100	100	65	71
≥ 2 predictors present	0.818	0.686–0.951	16	5	3	21	76	88	84	81	82
≥ 1 predictor present	0.542	0.372–0.711	21	0	22	2	100	8	49	100	51

M-CC LR-M with cholangiocarcinoma component, AUC area under the receiver operating characteristic curve, TP true positive, FN false negative, FP false positive, TN true negative, PPV positive predictive value, NPV negative predictive value

than evaluating agreement between two reviewers [14–19], agreement was measured between three independent reviewers, and this difference in design might also have introduced more prominent inter-rater variability.

While most literature attempting to differentiate histological subtypes of LR-M lesions have focused on imaging features exclusively, we found that integrating the serum tumor markers CA19-9 and AFP into a diagnostic model can significantly aid in this task. Similar to our results, CA19-9 has been reported to be associated with cholangiocarcinoma in previous studies. Prior work conducted on healthy populations and those with

chronic liver diseases and/or primary sclerosing cholangitis has found CA19-9 helpful in diagnosing cholangiocarcinoma, and the most commonly reported threshold values have been approximately 37 U/mL or 100 U/mL [22, 23, 28]. The CA19-9 threshold value derived in our study was 38 U/mL, relatively lower than that in published literature. This could be in part attributable to our specific patient population (LR-M patients with chronic hepatitis B virus infection). Another possible explanation is that other publications include perihilar and extrahepatic cholangiocarcinomas while LR-M lesions are typically peripheral, and the severe biliary obstruction often associated

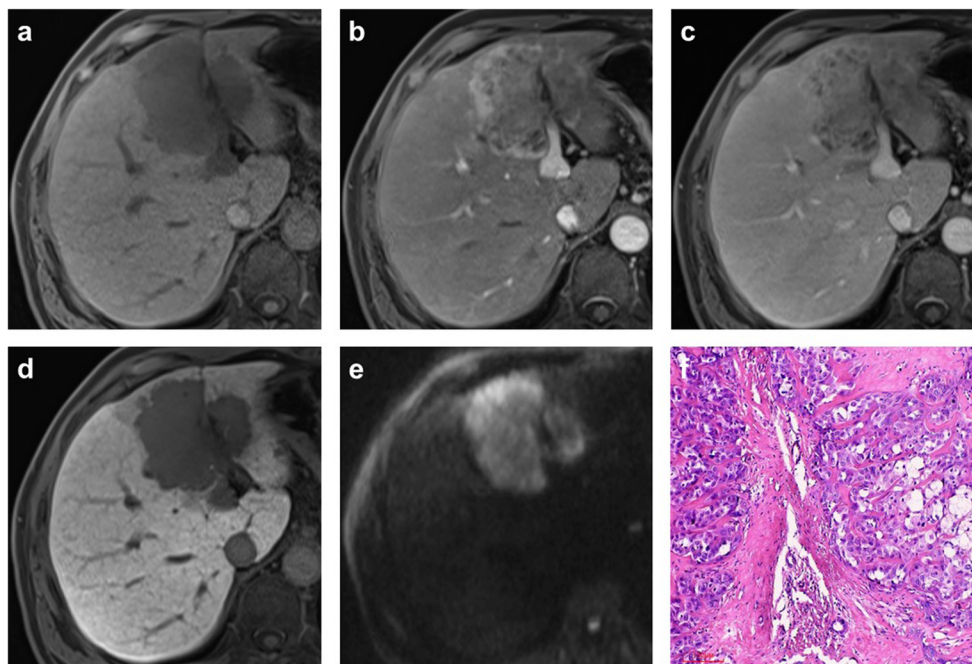


Fig. 3 Gadoxetate disodium-enhanced MR images of a 53-year-old male with chronic hepatitis B. Precontrast images (a) showed a 7.2-cm hypointense LR-M mass predominantly in segment IV with no “blood products in mass.” The mass demonstrated rim arterial phase hyperenhancement (b), mild central portal venous phase enhancement (c), hepatobiliary phase hypointensity (d), and non-targetoid restriction

on diffusion-weighted images (e, $b = 1200 \text{ s/mm}^2$). The patient’s serum CA19-9 and AFP levels were 8.46 U/mL and 3.44 ng/mL (one predictor present), respectively, so a *non-M-CC* diagnosis was established by Model F. The mass was histopathologically proven as hepatocellular carcinoma, with a representative section with hematoxylin-eosin staining at $\times 100$ magnification shown (f)

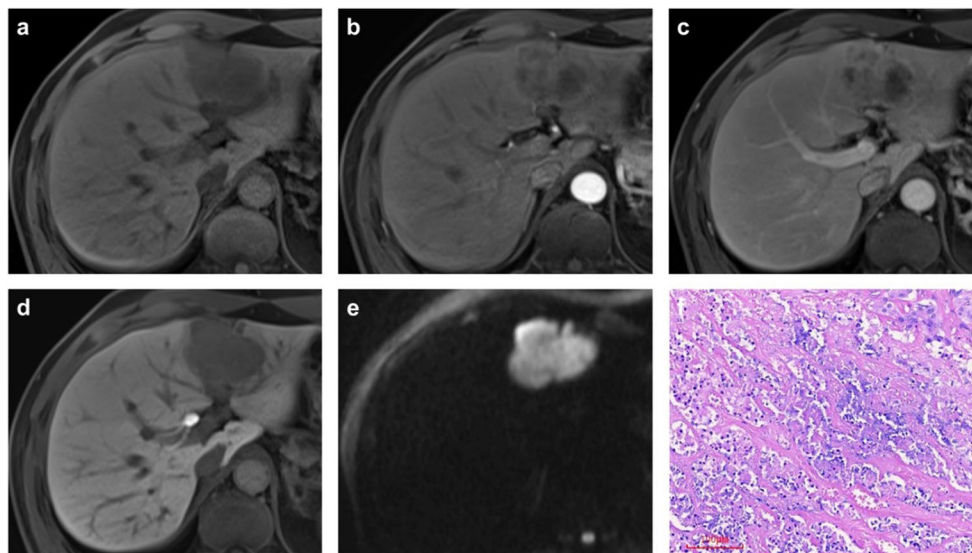


Fig. 4 Gadoxetate disodium-enhanced MR images of a 46-year-old male with chronic hepatitis B. Precontrast images (a) showed a 5.2-cm hypointense LR-M mass in segment II and IV with no “blood products in mass.” The mass demonstrated rim arterial phase hyperenhancement (b), central portal venous phase enhancement (c), hepatobiliary phase hypointensity (d), and non-targetoid restriction on diffusion-weighted

images (e, $b = 1200 \text{ s/mm}^2$). The patient’s serum CA19-9 and AFP levels were 42.65 U/mL and 46.07 ng/mL (three predictors present), respectively, so a *M-CC* diagnosis was established by Model F. The mass was histopathologically proven as intrahepatic cholangiocarcinoma, with a representative section with hematoxylin-eosin staining at $\times 100$ magnification shown (f)

with central lesions typically elevates CA19-9 values. Notably, the diagnostic role of CA19-9 in isolation remains a matter of debate due to overlap with other benign and malignant diseases [9]. However, because patients with LR-M lesions have a very high risk of malignancy relative to other populations [4], the use of CA19-9 in differentiation between *M-CC* and non-*M-CC* is more reliable.

As the most widely utilized tumor marker in patients with HCC, an elevated AFP level is a risk factor for HCC [13, 29, 30]. Interestingly, although AFP values greater than 8–200 ng/mL have been reported indicative of HCC in at-risk patients [18, 29, 31], we found that $\text{AFP} > 4.8 \text{ ng/mL}$ was a significant predictor for *M-CC*, rather than non-*M-CC*. One potential cause for this counterintuitive result was that *c-HCC-CCA*, a biphenotypic tumor frequently associated with elevation of both AFP and CA19-9 [6, 20, 32], was considered as *M-CC* in this study due to similar treatment strategy and prognosis to ICC. In the current study, patients with *c-HCC-CCA* had markedly higher AFP values compared with those with ICCs (Supplementary Material 2). However, this might have skewed our model and reduced the sensitivity for the ICC *M-CC*s. Therefore, the utility of AFP in discriminating between different LR-M entities remains unclear, and efforts in larger scale population to confirm or refute our finding should continue to be actively pursued.

In current treatment paradigms, optimal management of patients with LR-M lesions is often dictated by the presence or absence of a cholangiocarcinoma element within the tumor [6, 8, 9, 13, 29, 30]. Therefore, we focused our models on the detection of lesions with cholangiocarcinoma components

(including *c-HCC-CCA*) to refine the diagnosis of these challenging lesions. According to Model F, LR-M patients with all three criteria fulfilled should not undergo transplantation given 98% risk of *M-CC*, otherwise they may require biopsy to establish a definite diagnosis. When all three criteria are fulfilled, biopsy could be obviated in approximately 18% LR-M patients with excellent specificity. However, it should be noted that most included patients were beyond the Milan criteria [29]. Nevertheless, liver transplantation could still have been feasible after successful downstaging; hence, *M-CC* detection is clinically relevant for proper downstream management in this context.

Our study had several limitations. First, only 45 patients were included in this study. While this limited the statistical power of our study, LR-M lesions are relatively unusual, and it is challenging to collect larger cohorts in a single-center study. Additionally, although the sample was drawn from a consecutively enrolled prospective cohort, and feature reduction was performed before logistic regression analysis, there remains the possibility of selection bias and overfitting. Also, the relatively low prevalence of cirrhosis and high rate of hepatitis B virus infection of the enrolled patients might limit the application of our findings in a Western population. Second, all LI-RADS v2018 imaging features were evaluated based on EOB-MRI, but LI-RADS has been only minimally adapted for EOB-MRI because it was originally designed for MRI using extracellular contrast agents [33, 34]. Additionally, EOB is much less frequently used than extracellular contrast agents in Western countries. This might limit the generalizability of our model and highlighted the need for further assessment based on extracellular contrast agent-enhanced MRI.

Third, patients with concomitant LR-4 and/or LR-5 lesions were not included in the current study so as to concentrate our scope on patients who had LR-M lesions exclusively and to eliminate the confounding effects of LR-4/5 lesions on serum tumor markers. However, excluding these patients could have restricted our model generalizability. Fourth, in patients with multiple LR-M lesions, the largest lesion for each patient was selected for analysis and assumed to be the major driver of the tumor markers. However, we did not have a rational way to account for the effect of multiple lesions on these lab results. Finally, due to the single-center nature of this study, no external validation was available to test and refine our models. Therefore, although we found serum tumor markers valuable in our clinical setting, the findings still mandate further validation.

In conclusion, a model integrating CA19-9, AFP, and the LI-RADS feature “blood products in mass” achieved high diagnostic performance for cholangiocarcinoma-containing tumors (M-CCs) in at-risk patients with LR-M lesions. When all three criteria were fulfilled, the specificity for M-CC was 100%, which might reduce or eliminate the need for biopsy confirmation.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00330-020-07488-z>.

Funding This study has received funding by Research Grant of National Natural Science Foundation of China (No. 81771797).

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Hanyu Jiang.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry Two of the authors, Dong Xiao and Alaattin Erkanli, have significant statistical expertise.

Informed consent Written informed consent was not required for this study because we retrospectively analyzed data from a prospectively collected cohort (Clinical trial registration No. ChiCTR1900026668).

Ethical approval Institutional Review Board approval was obtained.

Study subjects or cohorts overlap In a previous study (Jiang H, Liu X, Chen J, et al (2019) Man or machine? Prospective comparison of the version 2018 EASL, LI-RADS criteria and a radiomics model to diagnose hepatocellular carcinoma. *Cancer Imaging* 19(1):84), we reported 30 patients included in the current study. While the previous work evaluated and compared the diagnostic accuracies of EASL v2018, LI-RADS v2018 criteria, and a radiomics model for HCC, the current study focused on the detection of M-CC in LR-M patients using a quite different methodology.

Methodology

- retrospective
- diagnostic or prognostic study
- performed at one institution

References

1. American College of Radiology (2018) CT/MRI LI-RADS version 2018. Available via <https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/LI-RADS/CT-MRI-LI-RADS-v2018>. Accessed November 1, 2019
2. Kierans AS, Makkar J, Guniganti P et al (2019) Validation of Liver Imaging Reporting and Data System 2017 (LI-RADS) criteria for imaging diagnosis of hepatocellular carcinoma. *J Magn Reson Imaging* 49(7):e205–e215
3. Lee SM, Lee JM, Ahn SJ et al (2019) LI-RADS version 2017 versus version 2018: diagnosis of hepatocellular carcinoma on gadoxetate disodium-enhanced MRI. *Radiology* 292(3):655–663
4. van der Pol CB, Lim CS, Sirlin CB et al (2019) 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* 156(4):976–986
5. Tang A, Bashir MR, Corwin MT et al (2018) Evidence supporting LI-RADS major features for CT- and MR imaging-based diagnosis of hepatocellular carcinoma: a systematic review. *Radiology* 286(1):29–48
6. Yin X, Zhang BH, Qiu SJ et al (2012) Combined hepatocellular carcinoma and cholangiocarcinoma: clinical features, treatment modalities, and prognosis. *Ann Surg Oncol* 19(9):2869–2876
7. Brunt E, Aishima S, Clavien PA et al (2018) cHCC-CCA: consensus terminology for primary liver carcinomas with both hepatocytic and cholangiocytic differentiation. *Hepatology* 68(1):113–126
8. Sapisochin G, Fidelman N, Roberts JP, Yao FY (2011) Mixed hepatocellular cholangiocarcinoma and intrahepatic cholangiocarcinoma in patients undergoing transplantation for hepatocellular carcinoma. *Liver Transpl* 17(8):934–942
9. Bridgewater J, Galle PR, Khan SA et al (2014) Guidelines for the diagnosis and management of intrahepatic cholangiocarcinoma. *J Hepatol* 60(6):1268–1289
10. Joo I, Lee JM, Yoon JH (2018) Imaging diagnosis of intrahepatic and perihilar cholangiocarcinoma: recent advances and challenges. *Radiology* 288(1):7–13
11. Clements O, Eliahoo J, Kim JU et al (2020) Risk factors for intrahepatic and extrahepatic cholangiocarcinoma: a systematic review and meta-analysis. *J Hepatol* 72(1):95–103
12. Kim JH, Joo I, Lee JM (2019) Atypical appearance of hepatocellular carcinoma and its mimickers: how to solve challenging cases using gadoxetic acid-enhanced liver magnetic resonance imaging. *Korean J Radiol* 20(7):1019–1041
13. Marrero JA, Kulik LM, Sirlin CB et al (2018) Diagnosis, staging, and management of hepatocellular carcinoma: 2018 Practice Guidance by the American Association for the Study of Liver Diseases. *Hepatology* 68(2):723–750
14. Kim YY, Kim MJ, Kim EH et al (2019) Hepatocellular carcinoma versus other hepatic malignancy in cirrhosis: performance of LI-RADS Version 2018. *Radiology* 291(1):72–80
15. Choi SH, Lee SS, Park SH et al (2019) LI-RADS classification and prognosis of primary liver cancers at gadoxetic acid-enhanced MRI. *Radiology* 290(2):388–397
16. Fraum TJ, Tsai R, Rohe E et al (2018) Differentiation of hepatocellular carcinoma from other hepatic malignancies in patients at risk: diagnostic performance of the Liver Imaging Reporting and Data System Version 2014. *Radiology* 286(1):158–172
17. Choi SH, Lee SS, Kim SY et al (2017) Intrahepatic cholangiocarcinoma in patients with cirrhosis: differentiation from hepatocellular carcinoma by using gadoxetic acid-enhanced MR imaging and dynamic CT. *Radiology* 282(3):771–781
18. Park HJ, Jang KM, Kang TW et al (2016) Identification of imaging predictors discriminating different primary liver tumours in patients

- with chronic liver disease on gadoxetic acid-enhanced MRI: a classification tree analysis. *Eur Radiol* 26(9):3102–3111
19. Ludwig DR, Fraum TJ, Cannella R et al (2019) Hepatocellular carcinoma (HCC) versus non-HCC: accuracy and reliability of Liver Imaging Reporting and Data System v2018. *Abdom Radiol (NY)* 44(6):2116–2132
 20. Horvat N, Nikolovski I, Long N et al (2018) Imaging features of hepatocellular carcinoma compared to intrahepatic cholangiocarcinoma and combined tumor on MRI using liver imaging and data system (LI-RADS) version 2014. *Abdom Radiol (NY)* 43(1):169–178
 21. Hwang J, Kim YK, Min JH et al (2017) Capsule, septum, and T2 hyperintense foci for differentiation between large hepatocellular carcinoma (≥ 5 cm) and intrahepatic cholangiocarcinoma on gadoxetic acid MRI. *Eur Radiol* 27(11):4581–4590
 22. Charatcharoenwitthaya P, Enders FB, Halling KC, Lindor KD (2008) Utility of serum tumor markers, imaging, and biliary cytology for detecting cholangiocarcinoma in primary sclerosing cholangitis. *Hepatology* 48(4):1106–1117
 23. Wang M, Gao Y, Feng H et al (2018) A nomogram incorporating six easily obtained parameters to discriminate intrahepatic cholangiocarcinoma and hepatocellular carcinoma. *Cancer Med* 7(3):646–654
 24. Chinese Society of Hepatology, Chinese Medical Association (2019) Chinese guidelines on the management of liver cirrhosis. *Zhonghua Gan Zang Bing Za Zhi* 27(11):846–865
 25. Bosman FT, Carneiro F, Hruban RH, Theise ND (2010) WHO classification of tumours of the digestive system. World Health Organization, Geneva
 26. Cerny M, Chernyak V, Olivie D et al (2018) LI-RADS version 2018 ancillary features at MRI. *Radiographics* 38(7):1973–2001
 27. Cerny M, Bergeron C, Billiard JS et al (2018) LI-RADS for MR imaging diagnosis of hepatocellular carcinoma: performance of major and ancillary features. *Radiology* 288(1):118–128
 28. Liang B, Zhong L, He Q et al (2015) Diagnostic accuracy of serum CA19-9 in patients with cholangiocarcinoma: a systematic review and meta-analysis. *Med Sci Monit* 21:3555–3563
 29. European Association for the Study of the Liver (2018) EASL clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol* 69(1):182–236
 30. Omata M, Cheng AL, Kokudo N et al (2017) Asia-Pacific clinical practice guidelines on the management of hepatocellular carcinoma: a 2017 update. *Hepatol Int* 11(4):317–370
 31. Marrero JA, Feng Z, Wang Y et al (2009) Alpha-fetoprotein, des-gamma carboxyprothrombin, and lectin-bound alpha-fetoprotein in early hepatocellular carcinoma. *Gastroenterology* 137(1):110–118
 32. Lee HS, Kim MJ, An C (2019) How to utilize LR-M features of the LI-RADS to improve the diagnosis of combined hepatocellular-cholangiocarcinoma on gadoxetate-enhanced MRI? *Eur Radiol* 29(5):2408–2416
 33. Elsayes KM, Fowler KJ, Chernyak V et al (2019) User and system pitfalls in liver imaging with LI-RADS. *J Magn Reson Imaging* 50(6):1673–1686
 34. Kim YY, Choi JY, Sirlin CB et al (2019) Pitfalls and problems to be solved in the diagnostic CT/MRI Liver Imaging Reporting and Data System (LI-RADS). *Eur Radiol* 29(3):1124–1132

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.