Title

Imaging Outcomes of Liver Imaging Reporting and Data System Version 2014 Category 2, 3, and 4 Observations Detected at CT and MR Imaging.

Journal

Radiology, 281(1)

ISSN

0033-8419

Authors

Tanabe, Masahiro
Kanki, Akihiko
Wolfson, Tanya
et al.

Publication Date

2016-10-01

DOI

10.1148/radiol.2016152173

Peer reviewed

Imaging Outcomes of Liver Imaging Reporting and Data System Version 2014 Category 2, 3, and 4 Observations Detected at CT and MR Imaging[1]

Masahiro Tanabe, MD[2]
Akihiko Kanki, MD[3]
Tanya Wolfson, MA
Eduardo A. C. Costa, MD
Adrija Mamidipalli, MD
Marilia P. F. D. Ferreira, MD
Cynthia Santillan, MD
Michael S. Middleton, MD, PhD
Anthony C. Gamst, PhD
Yuko Kono, MD
Alexander Kuo, MD
Claude B. Sirlin, MD

Purpose: To determine the proportion of untreated Liver Imaging Reporting and Data System (LI-RADS) version 2014 category 2, 3, and 4 observations that progress, remain stable, or decrease in category and to compare the cumulative incidence of progression in category.

Materials and Methods: In this retrospective, longitudinal, single-center, HIPAA-compliant, institutional review board–approved study, 157 patients (86 men and 71 women; mean age ± standard deviation, 59.0 years ± 9.7) underwent two or more multiphasic computed tomographic (CT) or magnetic resonance (MR) imaging examinations for hepatocellular carcinoma surveillance, with the first examination in 2011 or 2012. One radiologist reviewed baseline and follow-up CT and MR images (mean follow-up, 614 days). LI-RADS categories issued in the clinical reports by using version 1.0 or version 2013 were converted to version 2014 retrospectively; category modifications were verified with another radiologist. For index category LR-2, LR-3, and LR-4 observations, the proportions that progressed, remained stable, or decreased in category were calculated. Cumulative incidence curves for progression were compared according to baseline LI-RADS category (by using log-rank tests).

Results: All 63 index LR-2 observations remained stable or decreased in category. Among 166 index LR-3 observations, seven (4%) progressed to LR-5, and eight (5%) progressed to LR-4. Among 52 index LR-4 observations, 20 (38%) progressed to a malignant category. The cumulative incidence of progression to a malignant category was higher for index LR-4 observations than for index LR-3 or LR-2 observations (each $P < .001$) but was not different between LR-3 and LR-2 observations ($P = .155$). The cumulative incidence of progression to at least category LR-4 was trend-level higher for index LR-3 observations than for LR-2 observations ($P = .0502$).

Conclusion: Observations classified according to LI-RADS version 2014 categories are associated with different imaging outcomes.

Online supplemental material is available for this article.

[1] From the Liver Imaging Group, Department of Radiology (M.T., A.K., E.A.C.C., A.M., M.P.F.D.F., C.S., M.S.M., C.B.S.) and Division of Hepatology, Department of Medicine (Y.K., A.K.), University of California, San Diego, 408 Dickinson St, San Diego, CA 92103; and Computational and Applied Statistics Laboratory (CASL), SDSC–University of California, San Diego, La Jolla, Calif (T.W., A.C.G.). Received October 3, 2015; revision requested November 12; revision received January 18, 2016; accepted February 9; final version accepted February 26. **Address correspondence to** C.B.S. (e-mail: *csirlin@ucsd.edu*).

Current addresses:
[2] Department of Radiology, Yamaguchi University Graduate School of Medicine, Ube, Yamaguchi, Japan.
[3] Department of Radiology, Kawasaki Medical School, Kurashiki, Okayama, Japan.

Contrast material–enhanced computed tomography (CT) and magnetic resonance (MR) imaging are frequently used for the noninvasive diagnosis of hepatocellular carcinoma (HCC). Despite the important role of these modalities, until recently, there has been no standardized system for image interpretation and reporting (1).

With the Liver Imaging Reporting and Data System (LI-RADS), the American College of Radiology attempts to standardize the interpretation of CT and MR images and the reporting of findings in patients with cirrhosis or other risk factors for HCC (2,3). The system was first released in 2011 (LI-RADS version 1.0) and was updated in 2013 (LI-RADS version 2013) and 2014 (LI-RADS version 2014). As explained on the American College of Radiology LI-RADS Web site (2), categories are assigned to individual liver observations (lesions or pseudolesions) on the basis of the relative probability of being benign or malignant. Categories are assigned by using a combination of major features, ancillary features, and prior knowledge (2). The current version (LI-RADS version 2014) incorporates enhancement characteristics in the hepatobiliary phase by using hepatobiliary contrast agents.

LR-1 observations are those that are interpreted as definitely benign; this group includes cysts and typical hemangiomas. LR-5 observations are those with imaging features diagnostic of HCC, namely arterial phase hyperenhancement in conjunction with one or more additional major features (washout appearance, capsule appearance, or threshold growth), taking into account the observation diameter (2). The LR-5 criteria are intended to have near 100% specificity for the diagnosis of HCC. As defined in version 2014, these criteria are equivalent to those endorsed by the Organ Procurement and Transplantation Network for noninvasive diagnosis of HCC (4,5), and patients with LR-5 observations may be eligible for curative treatment, such as liver transplantation, in the absence of confirmatory biopsy. In addition, some LR-5 criteria, including the combination of arterial phase hyperenhancement and washout appearance, have been validated in prior studies (6–12). LR-M observations are those with features diagnostic for or highly suggestive of malignancy but in which the features are not specific for HCC (2).

Other observations are categorized as LR-2 (probably benign), LR-3 (intermediate probability for HCC), or LR-4 (probably HCC). The criteria for these categories were developed on the basis of expert opinion, and the outcomes of LR-2, LR-3, and LR-4 observations have not been studied extensively. In another retrospective single-center study, investigators evaluated the imaging outcomes of LR-3 observations and found that most LR-3 observations were hypervascular pseudolesions that remained stable or regressed (13). However, that study was limited by a small cohort size and lack of inclusion of LR-2 and LR-4 observations.

The purpose of this study was to determine, by using LI-RADS version 2014, the proportion of untreated LR-2, LR-3, and LR-4 observations that progress, remain stable, or decrease in category and to compare the cumulative incidence of progression to a higher category.

Advances in Knowledge

- Observations classified according to Liver Imaging Reporting and Data System version 2014 categories are associated with different imaging outcomes.
- Among 52 index LR-4 observations, 20 (38%) progressed to LR-5 (n = 19 observations) or LR-M (n = 1), 23 (44%) remained stable, and nine (17%) decreased in category.
- Among 166 index LR-3 observations, seven (4%) progressed to LR-5, eight (5%) progressed to LR-4, 38 (23%) remained stable, and 113 (68%) decreased in category.
- The cumulative incidence of progression to a malignant category (LR-5 or LR-M) was higher for index LR-4 observations than for index LR-3 or LR-2 observations ($P < .001$ for each comparison).

Implication for Patient Care

- LR-4 observations have a substantial risk of progression to a malignant category; depending on the diameter and clinical context, appropriate management considerations may include biopsy to confirm hepatocellular carcinoma or other malignancy, treatment without biopsy, or close imaging follow-up.

Materials and Methods

Study Design

This was a retrospective, observational, longitudinal, single-center study of patients who underwent clinical CT or MR imaging examinations for surveillance or diagnosis of HCC. Retrospective data collection and analysis were approved by our institutional review board, with waiver of written informed consent. The study was Health Insurance Portability and Accountability Act compliant.

Patient Selection

Our institution adopted LI-RADS when it was released in March 2011. Since

Published online before print
10.1148/radiol.2016152173 **Content codes:** CT GI MR

Radiology 2016; 281:129–139

Abbreviations:
CI = confidence interval
HCC = hepatocellular carcinoma
LI-RADS = Liver Imaging Reporting and Data System

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

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

Figure 1

Inclusion and Exclusion Criteria

Inclusion Criteria

1. At least one LR-2, LR-3, or LR-4 observation* reported on multiphasic CT or MR images obtained for HCC surveillance, diagnosis, or tumor response assessment between March 2011 and December 2012 (baseline examinations)
2. At least one additional multiphasic CT or MR imaging examination performed between March 2011 and March 2015 (follow-up examinations)

Exclusion Criteria

1. No follow-up examination performed at least 1 month after the baseline examination, unless the observation progressed to LR-5 within a month
2. Local-regional therapy of the observation performed without histologic assessment of the observation after the baseline examination and before the first follow-up examination
3. Surgical resection or liver transplantation performed without histologic assessment of the observation after the baseline examination and before the first follow-up examination

Figure 1: Chart provides the study inclusion and exclusion criteria. * = Based on the clinically reported LI-RADS version 1.0 category.

Figure 2

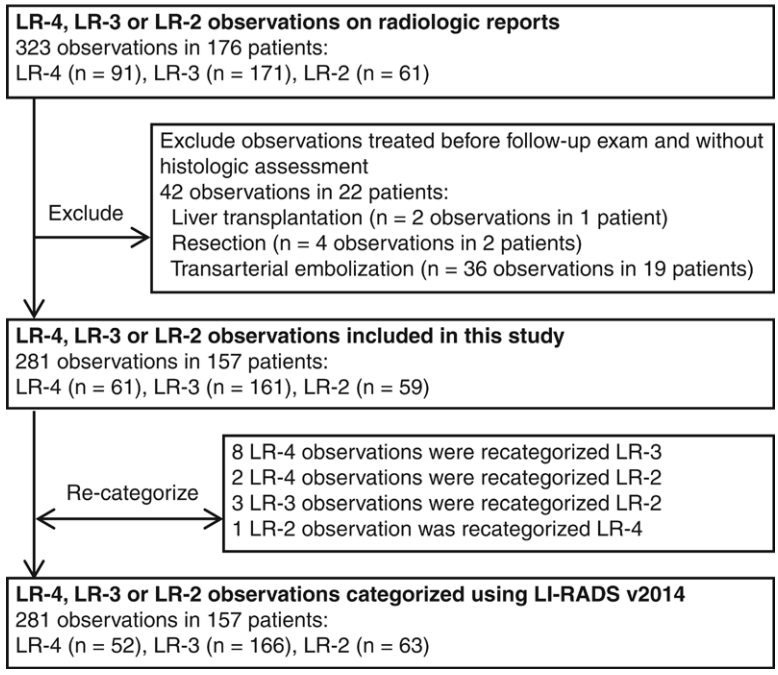

Figure 2: Flowchart illustrates the selection and LI-RADS version 2014 categorization of observations included in this study. Forty-two observations were excluded for treatment without histologic assessment before the first follow-up examination. Thirty-six lesions observed in 19 patients were embolized (29 were LR-4 observations targeted for embolization, and seven were LR-3 observations adjacent to HCC lesions targeted for embolization). Six observed lesions in three patients with HCC lesions elsewhere in the liver were treated surgically (two observed lesions in one patient were treated with transplantation, and four observed lesions in two patients were treated surgically), but these observations were not assessed histologically.

then, findings of all CT and MR imaging examinations performed for HCC surveillance, diagnosis, or follow-up have been reported for clinical care by using a standard template in which up to 10 individual observations per patient are given unique identifiers, assigned LI-RADS categories, and measured (long-axis diameter to the nearest millimeter). The observation identifiers are maintained in follow-up examinations, which permits longitudinal tracking, including evolution in LI-RADS category. The presence of tumor in veins (macrovascular invasion) is recorded (2).

We retrospectively searched the institutional radiology information systems to identify all consecutive patients with at least one LR-2, LR-3, or LR-4 observation reported on contrast-enhanced CT or MR images obtained from March 2011 through December 2012 and on images from at least one additional CT or MR imaging examination performed from March 2011 through March 2015. The first CT or MR imaging examination performed between March 2011 through December 2012 was considered the baseline examination; all subsequent examinations performed through March 2015 were considered follow-up examinations. Eligibility criteria are listed in Figure 1 and were applied to select the study cohort and observation set as illustrated in Figure 2. To reflect our entire experience with various LI-RADS categories at our institution and to reduce confirmation bias, we did not impose a minimum follow-up threshold. Of the 511 patients who underwent CT or MR imaging for HCC surveillance, diagnosis, or tumor response assessment from March 2011 through December 2012, 176 patients with 323 LR-2, LR-3, or LR-4 observations were identified. Forty-two observations in 22 patients were excluded because they were treated without histologic assessment after the baseline examination and before the first follow-up examination (Fig 2). The final study cohort and set of index observations are described in the Results section. Demographic, clinical, and pathology data were extracted from electronic medical records.

Imaging Techniques

As summarized in Figure 3, dynamic contrast-enhanced CT examinations were performed with 64– and 320–detector row scanners. MR imaging examinations were performed with 1.5-T and 3-T imaging units.

Figure 3

CT and MR Imaging Techniques

CT Examinations*

Axial CT images (120 kVp; 200–750 mAs, adjusted according to patient size; section thickness of 0.625 mm; table speed of 39.37 mm per rotation; and pitch of 0.987) were acquired in the following phases before and after contrast material injection:†

- Before contrast material administration
- Late hepatic arterial phase (timing with bolus-tracking software)
- Portal venous phase
- Delayed phase (3–4 minutes after contrast agent administration)

All images were reconstructed with a section thickness of 2.5–3.75 mm. Arterial and portal venous phase images were also reformatted in coronal and sagittal planes, with reconstructed section thickness of 3–4 mm.

MR Imaging Examinations‡

Coronal and axial single-shot T2-weighted imaging (nominal repetition time [msec], 1500; effective echo time [msec], 80–100; matrix, 256 × 192 for 1.5 T and 320 × 224 for 3 T; 6–8-mm section thickness; no parallel imaging for coronal acquisitions; acceleration factor of 1.25 to 2× for axial acquisitions)

Axial two-dimensional dual-gradient-echo in- and out-of-phase T1-weighted imaging (repetition time [msec], 150; flip angle, 70°; echo times [msec], 2.3 and 4.6 for 1.5 T and 2.3 and 5.8 for 3 T; matrix, 256 × 192 for 1.5 T and 320 × 224 for 3 T; 6–8-mm section thickness; and no parallel imaging)

Axial two-dimensional spin-echo echo-planar diffusion-weighted imaging (3000–4000/minimum; matrix, 100 × 128; 6–8-mm section thickness; and parallel imaging acceleration factor, 2×) with two *b* values as follows:

- Low (0–50 sec/mm^2) *b* value with 1–2 signals acquired
- Intermediate to high (300–800-sec/mm^2) *b* value with 4–8 signals acquired

Contrast-enhanced three-dimensional fat-suppressed dynamic T1-weighted imaging implemented as a volumetric interpolated breath-hold examination on the Siemens imaging unit and a liver acquisition with volume acceleration sequence on the GE imaging units (repetition time/echo time, minimum/minimum; flip angle, 15°; matrix, 256 × 128 to 288 × 192 for 1.5 T and 320 × 160 to 320 × 224 for 3 T; section thickness, 4–6 mm interpolated to have 50% overlap; and no parallel imaging performed, except as indicated) in the axial plane, except as indicated, and in the following phases before and after contrast material injection:§

- Before contrast material administration
- Late hepatic arterial phase, consisting of two separate consecutive acquisitions obtained within a single breath hold beginning 30 seconds after injection and using parallel imaging acceleration 2×
- Portal venous phase
- Delayed (axial and coronal) and transitional (axial) phases (3–5 minutes after contrast agent administration)
- Hepatobiliary phase (15–30 minutes after injection of gadoxetic acid) (axial and coronal)

Figure 3: Chart shows the typical imaging parameters used during the study period. Parameters may have varied slightly in individual patients. *CT examinations were performed with 64–detector row (Discovery CT750 HD; GE Medical Systems, Waukesha, Wis) and 320–detector row (Toshiba Aquilion; Toshiba America Medical Systems, Tustin, Calif) scanners. †For contrast-enhanced dynamic CT, ioxehol-350 was injected as a fixed volume of 125 mL at a rate of 4–5 mL/sec. ‡MR imaging examinations were performed with 1.5-T imaging units (Echospeed HD with an eight-channel coil, GE Medical Systems; or Siemens Symphony with a four-channel coil, Siemens Medical Systems, Erlangen, Germany) and a 3-T imaging unit (Signa Excite HD with an eight-channel coil; GE Medical Systems). For 3-T imaging, a dielectric pad was placed between the body wall and the torso phased-array coil. §For contrast-enhanced dynamic MR imaging, gadoxetic acid (Eovist, Bayer-Schering Pharma, Berlin, Germany; 0.025 mmol of gadolinium per kilogram of body weight), gadobenate dimeglumine (MultiHance, Bracco Diagnostics, Princeton, NJ; 0.1 mmol of gadolinium per kilogram of body weight), or gadobutrol (Gadavist, Bayer Pharma; 0.05 mmol of gadolinium per kilogram of body weight) was injected at 1 mL/sec (gadoxetic acid, gadobutrol) or 2 mL/sec (gadobenate dimeglumine), followed by a 40-mL saline flush at 2 mL/sec.

LI-RADS Categorization

The LI-RADS category and presence or absence of tumor in veins for all index observations and examinations were reported clinically by one of nine academic abdominal radiologists at our center, each with a minimum of 4 years of postfellowship experience in abdominal imaging. LI-RADS version 1.0 was used in radiology reports for examinations performed from March 2011 through December 2012. A modified version of LI-RADS version 2013 was used in radiology reports for examinations performed between January 2013 and November 2014, and version 2014 was used after November 2014.

LI-RADS categories issued in the clinical reports by using version 1.0 or modified version 2013 were subsequently converted to version 2014 by two academic abdominal radiologists not involved in the clinical reporting who were blinded to clinical and pathologic results, as well as the imaging outcomes (M.T. [reader 1] and A.K. [reader 2], with 11 and 10 years of postfellowship experience in abdominal imaging, respectively). First, reader 1 reviewed the radiology reports from the baseline multiphasic CT or MR imaging examinations and from all follow-up CT and MR imaging examinations until the observation progressed to a malignant category (LR-5 or LR-M) or, for observations that did not progress to a malignant category, until the observation was treated or lost to follow-up. The clinically reported category for each index observation was recorded. Additionally, observation diameter, location (left or right lobe), presence of any LR-5 observations elsewhere in the liver, and history of prior HCC treatment were recorded at baseline. Since index LR-4 observations were thought a priori to have the greatest risk of progression, they were reviewed in greater detail, and the diameters were recorded at each follow-up time point, not just at baseline. Second, reviewer 1 reviewed the images from the baseline examinations, the follow-up examinations in which the clinically reported category was different than on the antecedent

examination, and the final examinations. For examination findings reported with version 1.0 or 2013, reader 1 retrospectively converted each category to version 2014 categories, and for examination findings reported with version 2014, reader 1 retrospectively confirmed or corrected the reported version 2014 categories. For baseline and final examinations, modifications in category from the clinical reports were verified by reader 2. To do this, reader 2 reviewed the modifications made by reader 1 without reader 1 being present. Reader 2 agreed with and accepted all of reader 1's modifications but one; reader 1 and reader 2 then reviewed this case together, decided that the clinically reported LI-RADS category was correct, and, in consensus, rejected the modification. Thus, all baseline and final version 2014 category codes used in the analysis were assigned in consensus (either by reader 1 and the clinical report if there was agreement with the report or by reader 1 and reader 2 if there was disagreement).

Statistical Analysis

Statistical analyses were performed with R version 3.0.2 (R Foundation for Statistical Computing, Vienna, Austria, 2013). Analyses were conducted at the observation level.

Cohort and observation characteristics were summarized descriptively. Follow-up data were summarized, overall and according to baseline LI-RADS category.

Cumulative incidence curves for progression to malignant LI-RADS category (LR-5 or LR-M according to imaging findings) were generated separately for observations categorized at baseline as LR-2, LR-3, or LR-4. Cumulative incidence curves for progression to at least category LR-4 (ie, to LR-4, LR-5, or LR-M according to imaging findings) were generated separately for observations categorized at baseline as LR-2 or LR-3. In generating these curves, we used only imaging-based LI-RADS categories to assess progression. Although pathology results were recorded, we did not adjust or confirm the final category on the basis of histology data, as these were infrequently available. Curves were compared pairwise by using log-rank tests with the resampling extension to adjust for the variable number of observations per subject. At each resampling iteration, one observation per patient was selected at random; test statistics were averaged over the iterations, and average log-rank test *P* values were computed. Since there were three pairwise comparisons (LR-2 vs LR-3, LR-3 vs LR-4, and LR-2 vs LR-4) for the analysis of progression to a malignant category, a Bonferroni-adjusted α level of 0.05/3 was used as a significance criterion for individual tests.

Figure 4

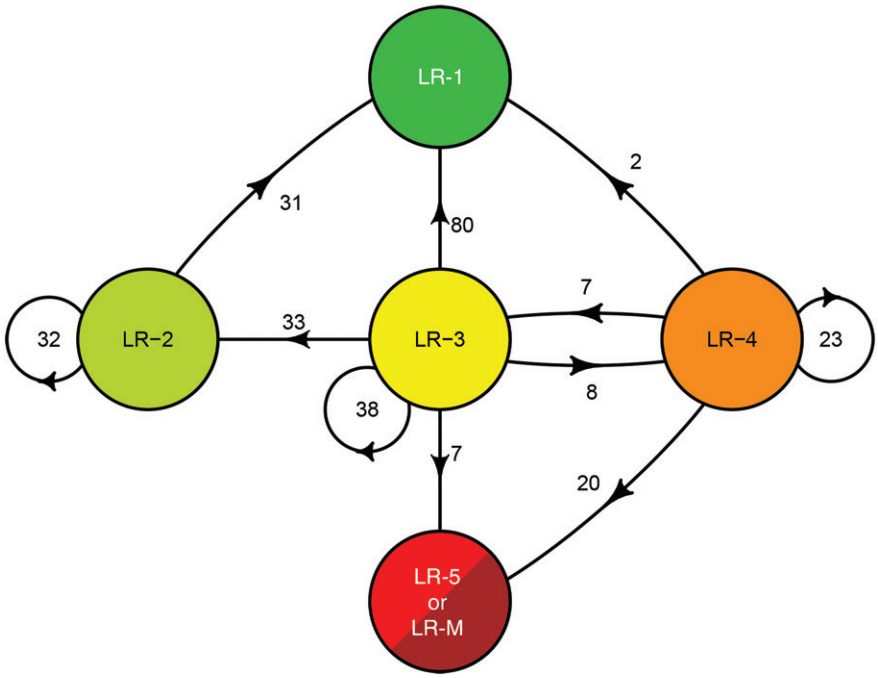

Figure 4: Diagram illustrates the transitions in version 2014 categories during follow-up (from baseline to final examination) for all observations included in this study. Data are numbers of observations. Circles are filled by using the LI-RADS version 2014 color codes. Nineteen index LR-4 observations progressed to LR-5, and one progressed to LR-M. Seven index LR-3 observations progressed to LR-5, and none progressed to LR-M.

Results

Study Cohort

The final study cohort comprised 157 patients (mean age ± standard deviation, 59.0 years ± 9.7 [range, 32–95 years]; including 86 men [mean age, 58 years ± 9.7; range, 39–95 years] and 71 women [mean age, 61 years ± 9.4; range, 32–81 years]). All patients had chronic liver disease, and 155 (98.7%) had cirrhosis. Ninety-eight of 157 patients (62.4%) had hepatitis C virus infection, 16 (10.2%) had hepatitis B virus infection, 18 (11.5%) had alcoholic liver disease, 11 (7.0%) had nonalcoholic steatohepatitis, two (1.3%) had autoimmune hepatitis, two (1.2%) had primary biliary cirrhosis, nine (5.7%) had cryptogenic cirrhosis, and one (0.6%) had both hepatitis B and C virus infections. These patients had a total of 281 index observations.

As illustrated in Figure 2, baseline categories of 14 of the 281 observations (5.0%) were modified (see Appendix E1 [online]) after retrospective image review and conversion to version 2014 categories. After these modifications, the final distribution of version 2014

Follow-up Summary for Index Observations, Stratified according to Baseline and Final LI-RADS Categories by Using Version 2014

Baseline Category	Final Category of LR-1	Final Category of LR-2	Final Category of LR-3	Final Category of LR-4	Final Category of LR-5 or LR-M
LR-2					
No. of observations	31	32	0	0	0
Mean follow-up (d)	802	549	…	…	…
Median follow-up (d)	890	493	…	…	…
Follow-up range (d)	129–1344	159–1315	…	…	…
Follow-up interquartile range (d)	548–982	258–722	…	…	…
LR-3					
No. of observations	80	33	38	8	7
Mean follow-up (d)	799	775	469	386	424
Median follow-up (d)	845	858	289	356	422
Follow-up range (d)	141–1377	161–1352	126–1232	129–746	200–605
Follow-up interquartile range (d)	385–1140	383–1117	182–674	192–496	340–531
LR-4					
No. of observations	2	0	7	23	20
Mean follow-up (d)	850	…	663	333	210
Median follow-up (d)	850	…	541	212	175
Follow-up range (d)	537–1162	…	71–1305	91–1230	22–557
Follow-up interquartile range (d)	693–1006	…	275–1086	157–480	128–201

categories at baseline was 52 LR-4 observations, 166 LR-3 observations, and 63 LR-2 observations. All patients had at least one follow-up CT or MR imaging examination (mean number of follow-up examinations, 3.9 [range, 1–13 examinations]; mean duration of total follow-up, 614 days [median, 538 days; range, 22–1377 days; interquartile range, 261–969 days]). Baseline characteristics of index observations and follow-up statistics are provided in Appendix E1 (online).

Longitudinal Follow-up of Index LR-2, LR-3, and LR-4 Observations

Transitions between baseline and final LI-RADS categories are illustrated in Figure 4. Follow-up durations are summarized in Table 1.

Outcome of index LR-2 observations.—Among 63 index LR-2 observations, none progressed, 32 (51%) remained stable, and 31 (49%) decreased in category. None were assessed histologically. Thus, 0 of 63 observations (0%; 95% confidence interval [CI]: 0%, 5.7%) progressed to a higher category.

Outcome of index LR-3 observations.—Among 166 index LR-3 observations, seven (4%) progressed to LR-5 (Fig 5) (no observations progressed within 6 months, two observations progressed between 6 and 12 months, and five observations progressed at more than 12 months), eight (5%) progressed to LR-4 (no observation progressed within 3 months, two observations progressed between 3 and 6 months, three observations progressed between 6 and 12 months, and three observations progressed at more than 12 months), 38 (23%) remained stable, and 113 (68%) decreased in category. Thus, seven of 166 LR-3 observations (4%; 95% CI: 1.7%, 8.5%) progressed to LR-5, and 15 of 166 observations (9%; 95% CI: 5.1%, 14.5%) progressed to LR-4 or LR-5. No index LR-3 observation progressed to LR-M. At the time of documented progression to LR-5, the seven index LR-3 observations had grown from a mean diameter of 11 mm (range, 4–15 mm) to a mean diameter of 22 mm (range, 16–30 mm). None showed imaging evidence of tumor in veins. None were assessed histologically.

Outcome of index LR-4 observations.—The time course and final outcome for each index LR-4 observation are summarized in Figure 6. Among 52 index LR-4 observations, 20 (38%) progressed to LR-5 (n = 19) or LR-M (n = 1) during follow-up (four observations progressed within 3 months, 11 observations progressed between 3 and 6 months, one observation progressed between 6 and 12 months, and four observations progressed at more than 12 months), 23 (44%) remained stable, and nine (17%) decreased in category. Thus, 20 of 52 LR-4 observations (38.5%; 95% CI: 25.3%, 53%) progressed to LR-5, and 21 of 52 observations (40%; 95% CI: 27%, 54.9%) progressed to LR-5 or LR-M. Follow-up data are summarized in Table 1; additional details are provided in Appendix E1 (online).

Cumulative incidence of progression.—As shown in Figure 7, the cumulative incidence of progression to a malignant category (LR-5 or LR-M according to imaging findings) was higher for index LR-4 observations than for index LR-3 or LR-2 observations (each $P < .001$). The cumulative incidence was not higher, however, for index LR-3 observations than for index LR-2 observations ($P = .155$). As shown in Figure 8, the cumulative incidence of progression to at least category LR-4 (according to imaging findings) was higher with borderline statistical significance ($P = .0502$) for index LR-3 observations than for index LR-2 observations.

Discussion

In this single-center retrospective study, LR-2, LR-3, and LR-4 observations had different imaging outcomes. By using version 2014, 38% of index LR-4 observations progressed to a malignant category (LR-5 or LR-M)—usually within 6 months and sometimes within 3 months. As described in Appendix E1 (online), one index LR-4 observation progressed to LR-5V, and three index LR-4 observations progressed to LR-5

Figure 5

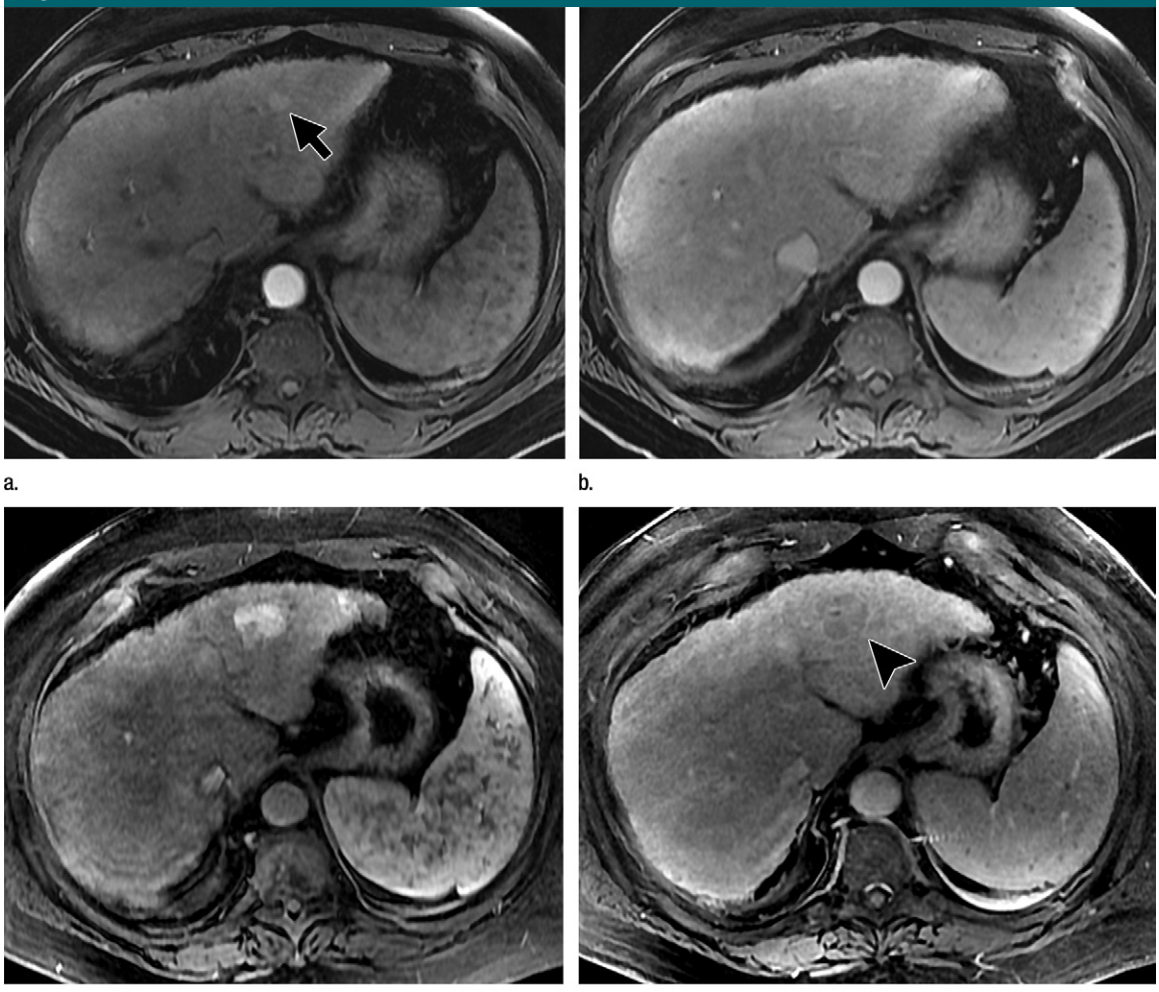

Figure 5: Axial MR images obtained in a 56-year-old man with an LR-3 observation that progressed to LR-5 (according to LI-RADS version 2014). **(a)** Arterial phase image acquired after administration of gadobenate dimeglumine shows a small (13-mm) nodular hyperenhancing observation in the left lobe (arrow). **(b)** Observation shows an isointensity in the portal venous phase that is occult on images obtained with other sequences (not shown). **(c)** Nineteen months later without any interim follow-up examinations, the observation has grown to 33 mm in diameter, continues to show arterial phase hyperenhancement, and **(d)** now shows partial hypointensity (arrowhead) relative to liver parenchyma in the portal venous phase. Due to the combination of imaging features (diameter $\geq$ 20 mm, arterial phase hyperenhancement, partial washout appearance, and threshold growth), the observation now is categorized as LR-5. Notice that at 19 months, the mass has become heterogeneous, with some areas enhancing more than others.

or LR-M, exceeding 50 mm in diameter. Of the 23 that remained stable in category, 43% grew by at least 3 mm during follow-up, 48% underwent local-regional treatment despite category stability, and 57% grew and/or were treated. Of the nine that decreased in category, six can reasonably be interpreted as nonmalignant on the basis of spontaneous disappearance, meaningful diameter reduction, or more than 2-year follow-up, while the other three had insufficient follow-up to exclude slow-growing malignancy. By comparison, only 4% of index LR-3 observations progressed to LR-5 (none within 6 months), and 7%–9% progressed to either LR-4 or LR-5; most remained stable or decreased in category. No LR-2 observations progressed. As expected, LR-4 observations had the highest cumulative incidence of progression to a malignant category. When compared with index LR-2 observations, index LR-3 observations had trendwise higher cumulative incidence of progression to at least category LR-4.

These findings have important management implications for institutions that use LI-RADS. LR-4 observations have substantial risk of progression to LR-5 or LR-M, with some advancing to cancers outside Milan criteria (14). However,

Figure 6

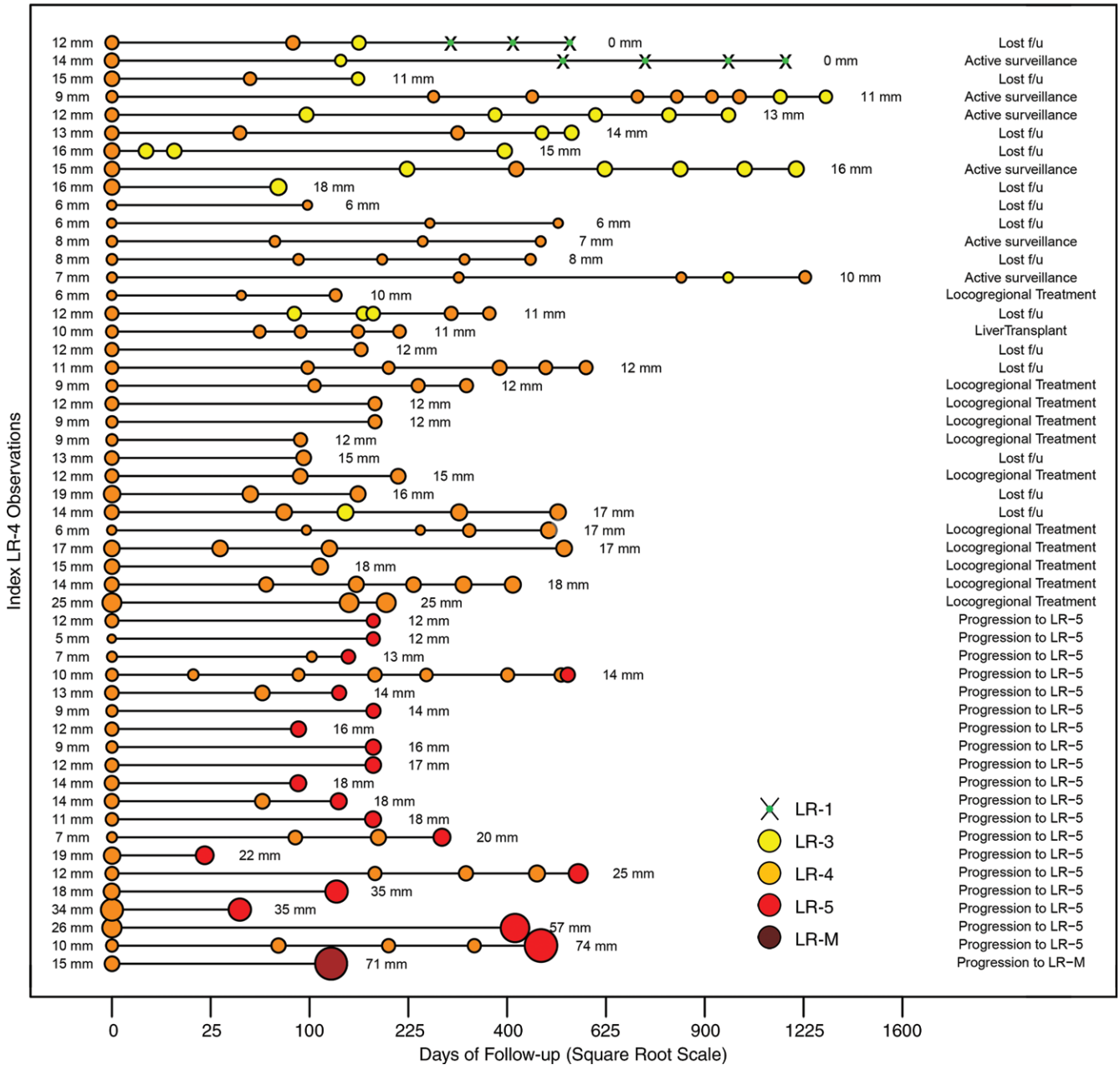

Figure 6: Graph shows the time course and final outcome for each index LR-4 observation (according to LI-RADS version 2014). The baseline and final diameter, the category at each time point, and the outcome for each observation are shown. The circles at each time point are proportional to the square root of the diameter of the observation. The circles are filled by using the LI-RADS version 2014 color codes (see the embedded legend). × = The observation was no longer visible at the corresponding time point (ie, spontaneous disappearance); observations that spontaneously disappear are categorized as LR-1 in LI-RADS version 2014. *f/u* = follow-up.

the rate and degree of progression are variable. Consequently, the optimal management of LR-4 observations is not straightforward. Depending on clinical and other considerations, reasonable options may include close imaging follow-up, biopsy, other diagnostic tests, or treatment without biopsy confirmation. If imaging follow-up is selected, our findings suggest that the time interval should be no more than 3 months, since progression to a malignant category can be rapid (75% of those that progressed to LR-5 or LR-M did so within 6 months, and 20% did so within 3 months).

Figure 7

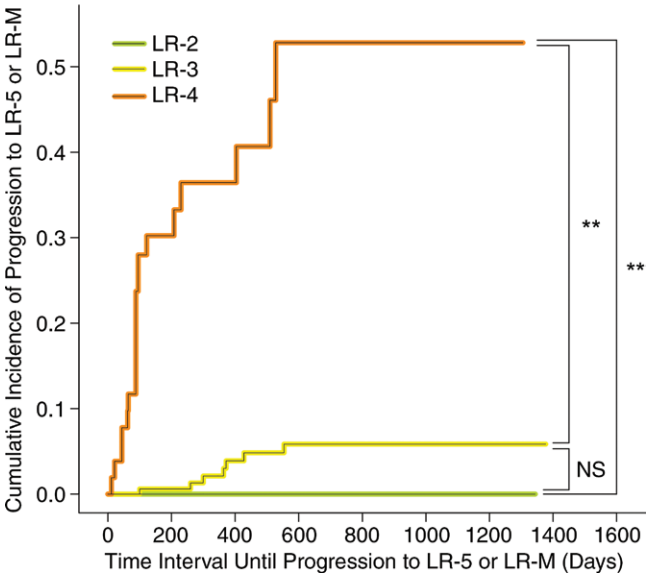

Figure 7: Graph shows the cumulative incidence of progression to a malignant category (LR-5 or LR-M) for index LR-2, LR-3, and LR-4 observations (according to LI-RADS version 2014). Curves show the cumulative incidence of progression to a malignant category (LR-5 or LR-M) of observations categorized as LR-2, LR-3, or LR-4 at baseline. Curves were compared by using average log-rank tests. In these tests, the χ^2 statistics were averaged over multiple iterations, and average *P* values were obtained; for each patient with at least two observations, one observation was selected at random in each iteration. *NS* = not significant. ** = $P < .001$.

Figure 8

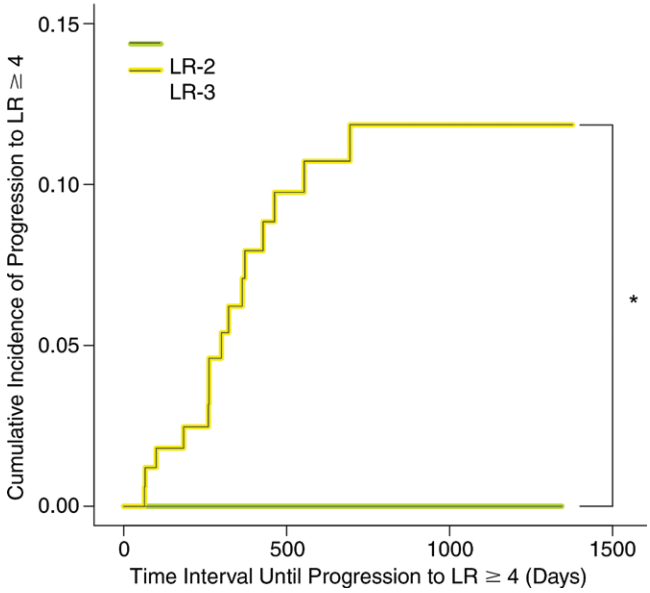

Figure 8: Graph shows the cumulative incidence of progression to at least category LR-4 for index LR-2 and LR-3 observations (according to LI-RADS version 2014). Curves show the cumulative incidence of progression to at least category LR-4 of observations categorized as LR-2 or LR-3 at baseline. Curves were compared by using an average log-rank test. In this test, the χ^2 statistic was averaged over multiple iterations, and an average *P* value was obtained; for each patient with at least two observations, one observation was selected at random in each iteration. * = $P = .0502$.

Moreover, follow-up should probably be conducted with CT or MR imaging rather than ultrasonography (US) to ensure that the same lesion or lesions are monitored and that changes in enhancement characteristics, especially those relevant to LR-5 categorization (arterial phase enhancement, washout appearance, and capsule appearance), can be identified. Following up LR-4 observations, however, may have risks. Not all patients can adhere to follow-up recommendations and, as illustrated by the LR-4 observation that progressed to LR-5V, initial stability does not exclude future rapid growth and aggressive behavior. Another complication is that LR-4 observations do not meet criteria for Organ Procurement and Transplantation Network class 5 (4,5) and so do not provide priority points for liver transplantation. Consequently, there may be reluctance to treat LR-4 observations in liver transplant candidates unless priority can be assigned on the basis of LR-5 observations elsewhere in the liver.

By comparison, less frequent imaging follow-up, perhaps every 6 months, probably suffices for LR-2 and LR-3 observations. In our study, these had low progression risk, and there were no recorded instances of progression to LR-5V, to LR-5 exceeding 50 mm, or to LR-M. Lack of progression does not prove benignity of these observations, however, since the total follow-up duration for many observations was insufficient to exclude slowly growing neoplasms. While our study suggests that a follow-up interval of 6 months may be reasonable, the study was not designed to determine the imaging modality that should be used for follow-up. Current clinical practice guidelines recommend US for HCC surveillance (15–18), but many LR-2 and LR-3 observations detected with CT and MR imaging are likely to be undetectable sonographically. Whether patients with LR-2 and LR-3 observations detected with CT and MR imaging should undergo surveillance with CT or MR imaging rather than US requires further study.

Our findings with regard to LR-3 observations are in keeping with those of Choi et al, who reported that 94% of LR-3 observations identified at gadoxetic acid–enhanced MR imaging remained stable or decreased in category during imaging follow-up (13). In no prior study have investigators examined the imaging outcome of LR-2 or LR-4 observations, to our knowledge.

The different imaging outcomes of LR-2, LR-3, and LR-4 observations provide preliminary validation of these categories, which were developed mainly on the basis of expert opinion. Partial validation also is provided by Darnell and colleagues, who showed that 96% of LR-4 observations with a histologic

reference standard were HCC (19). The LR-4 ("probably HCC") category is intended to convey high probability of HCC, and this was confirmed by these investigators. Nevertheless, some refinement of LI-RADS categorization may be needed. In our study, most LR-3 observations did not progress. Future refinement of LI-RADS may be needed to permit categorization as LR-2 of at least some observations that, in the current system, are categorized LR-3. Also, as mentioned earlier, LR-4 observations had variable outcomes; research is needed to identify and validate imaging features that better predict their outcomes.

This study had limitations. Because of its retrospective nature, numerous factors varied according to patient and observation, including imaging modality, imaging technique, and, as described in Appendix E1 (online), the frequency and duration of follow-up. Since imaging techniques were not standardized, some transitions between categories may have reflected differences in technique rather than true transitions. Future studies of LI-RADS outcomes would benefit from prospective design and standardized imaging technique and follow-up interval. Our study was performed at a single center, which limits generalizability, and had only a modest number (n = 52) of LR-4 observations. Larger, multicenter studies are needed to confirm and expand our results. Observations were recategorized by using version 2014 with knowledge of the reported categories and in consensus with a second radiologist. This does not reflect actual clinical practice and perhaps provides an idealized assessment; future work is needed to track the outcomes of LI-RADS observations as reported clinically. Although modifications in baseline and final categories were verified by a second radiologist, modifications in interim categories were not. We did not assess interreader agreement for LI-RADS categorization. Since other investigators have suggested that interreader agreement for LI-RADS imaging features (13) and for LI-RADS categories (20) may be modest, future studies should include independent reviews by multiple radiologists. Our study did not address how ancillary features affect LI-RADS categorization, as this was beyond the study scope. Because biopsy of nodules suspicious for malignancy is rarely performed at our institution, an unavoidable limitation was that most observations were unconfirmed pathologically. Finally, some observations were lost to follow-up before the outcome could be established reliably.

In conclusion, LR-2, LR-3, and LR-4 observations have different imaging outcomes. About two-fifths of LR-4 observations progressed to a malignant category; three-quarters that progressed did so within 6 months. Of those that did not progress in category, more than two-fifths grew during follow-up, and almost half were treated despite category stability. Most LR-3 and all LR-2 observations remained stable or decreased in category. These different imaging outcomes provide preliminary validation for categories that were developed on the basis of expert opinion. However, as this was a single-center retrospective study, our results should be interpreted as preliminary rather than definitive. Prospective multicenter studies are needed to validate our results, further refine the LI-RADS categories, and collect the data to inform optimal management strategies.

Disclosures of Conflicts of Interest: M.T. disclosed no relevant relationships. **A.K.** disclosed no relevant relationships. **T.W.** disclosed no relevant relationships. **E.A.C.C.** disclosed no relevant relationships. **A.M.** disclosed no relevant relationships. **M.P.F.D.F.** disclosed no relevant relationships. **C.S.** disclosed no relevant relationships. **M.S.M.** disclosed no relevant relationships. **A.C.G.** disclosed no relevant relationships. **Y.K.** disclosed no relevant relationships. **A.K.** disclosed no relevant relationships. **C.B.S.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: author received payment from VirtualScopics for consultancy; institution received grants from General Electric Healthcare, Pfizer, Siemens, and Guerbet. Other relationships: disclosed no relevant relationships.

References

1. Tang A, Cruite I, Sirlin CB. Toward a standardized system for hepatocellular carcinoma diagnosis using computed tomography and MRI. Expert Rev Gastroenterol Hepatol 2013;7(3):269–279.
2. American College of Radiology. Liver Imaging Reporting and Data System. http://www.acr.org/Quality-Safety/Resources/LI-RADS. Accessed April 12, 2016.
3. Purysko AS, Remer EM, Coppa CP, Leão Filho HM, Thupili CR, Veniero JC. LI-RADS: a case-based review of the new categorization of liver findings in patients with end-stage liver disease. RadioGraphics 2012;32(7):1977–1995.
4. Wald C, Russo MW, Heimbach JK, Hussain HK, Pomfret EA, Bruix J. New OPTN/UNOS policy for liver transplant allocation: standardization of liver imaging, diagnosis, classification, and reporting of hepatocellular carcinoma. Radiology 2013;266(2):376–382.
5. Pomfret EA, Washburn K, Wald C, et al. Report of a national conference on liver allocation in patients with hepatocellular carcinoma in the United States. Liver Transpl 2010;16(3):262–278.
6. Sersté T, Barrau V, Ozenne V, et al. Accuracy and disagreement of computed tomography and magnetic resonance imaging for the diagnosis of small hepatocellular carcinoma and dysplastic nodules: role of biopsy. Hepatology 2012;55(3):800–806.
7. Sharma P, Kalb B, Kitajima HD, et al. Optimization of single injection liver arterial phase gadolinium enhanced MRI using bolus track real-time imaging. J Magn Reson Imaging 2011;33(1):110–118.
8. Hecht EM, Holland AE, Israel GM, et al. Hepatocellular carcinoma in the cirrhotic liver: gadolinium-enhanced 3D T1-weighted MR imaging as a stand-alone sequence for diagnosis. Radiology 2006;239(2):438–447.
9. Marrero JA, Hussain HK, Nghiem HV, Umar R, Fontana RJ, Lok AS. Improving the prediction of hepatocellular carcinoma in cirrhotic patients with an arterially-enhancing liver mass. Liver Transpl 2005;11(3):281–289.
10. Carlos RC, Kim HM, Hussain HK, Francis IR, Nghiem HV, Fendrick AM. Developing a prediction rule to assess hepatic malignancy in patients with cirrhosis. AJR Am J Roentgenol 2003;180(4):893–900.
11. Khan AS, Hussain HK, Johnson TD, Weadock WJ, Pelletier SJ, Marrero JA. Value of delayed hypointensity and delayed enhancing rim in magnetic resonance imaging diagnosis of small hepatocellular carcinoma in the cirrhotic liver. J Magn Reson Imaging 2010;32(2):360–366.
12. Forner A, Vilana R, Ayuso C, et al. Diagnosis of hepatic nodules 20 mm or smaller in cir-

rhosis: prospective validation of the noninvasive diagnostic criteria for hepatocellular carcinoma. Hepatology 2008;47(1):97–104.

13. Choi JY, Cho HC, Sun M, Kim HC, Sirlin CB. Indeterminate observations (liver imaging reporting and data system category 3) on MRI in the cirrhotic liver: fate and clinical implications. AJR Am J Roentgenol 2013; 201(5):993–1001.
14. Mazzaferro V, Regalia E, Doci R, et al. Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. N Engl J Med 1996;334(11):693–699.
15. Tan CH, Low SC, Thng CH. APASL and AASLD consensus guidelines on imaging diagnosis of hepatocellular carcinoma: a review. Int J Hepatol 2011;2011:519783.
16. Kudo M, Izumi N, Kokudo N, et al. Management of hepatocellular carcinoma in Japan: consensus-based clinical practice guidelines proposed by the Japan Society of Hepatology (JSH) 2010 updated version. Dig Dis 2011;29(3):339–364.
17. Bruix J, Sherman M; American Association for the Study of Liver Diseases. Management of hepatocellular carcinoma: an update. Hepatology 2011;53(3):1020–1022.
18. European Association for the Study of the Liver; European Organisation for Research and Treatment of Cancer. EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. J Hepatol 2012; 56(4):908–943. [Published correction appears in J Hepatol 2012;56(6):1430.]
19. Darnell A, Forner A, Rimola J, et al. Liver imaging reporting and data system with MR imaging: evaluation in nodules 20 mm or smaller detected in cirrhosis at screening US. Radiology 2015;275(3):698–707.
20. Davenport MS, Khalatbari S, Liu PS, et al. Repeatability of diagnostic features and scoring systems for hepatocellular carcinoma by using MR imaging. Radiology 2014;272(1): 132–142.