HEPATOBILIARY-PANCREAS

Check for updates

Longitudinal evolution of CT and MRI LI-RADS v2014 category 1, 2, 3, and 4 observations

Cheng William Hong[1] (ORCID) · Charlie C. Park[2] · Adrija Mamidipalli[1] · Jonathan C. Hooker[1] · Soudabeh Fazeli Dehkordy[1] · Saya Igarashi[1] · Mohanad Alhumayed[1] · Yuko Kono[3] · Rohit Loomba[3] · Tanya Wolfson[4] · Anthony Gamst[4] · Paul Murphy[1] · Claude B. Sirlin[1]

Abstract

Objectives This study assesses the risk of progression of Liver Imaging Reporting and Data System (LI-RADS) categories, and the effects of inter-exam changes in modality or radiologist on LI-RADS categorization.

Methods Clinical LI-RADS v2014 CT and MRI exams at our institution between January 2014 and September 2017 were retrospectively identified. Untreated LR-1, LR-2, LR-3, and LR-4 observations with at least one follow-up exam were included. Three hundred and seventy-two observations in 214 patients (149 male, 65 female, mean age 61 ± 10 years) were included during the study period (715 exams total). Cumulative incidence curves for progression to malignant LI-RADS categories (LR-5 or LR-M) and to LR-4 or higher were generated for each index category and compared using log-rank tests with a resampling extension. Relationships between inter-exam changes in LI-RADS category and modality or radiologist, adjusted for inter-exam time intervals, were modeled using mixed effect logistic regressions.

Results Median inter-exam follow-up interval and total follow-up duration were 123 and 227 days, respectively. Index LR-1, LR-2, LR-3, and LR-4 differed significantly in their cumulative incidences of progression to malignant categories ($p < 0.0001$), which were 0%, 2%, 7%, and 32% at 6 months, respectively. Index LR-1, LR-2, and LR-3 differed significantly in cumulative incidences of progression to LR-4 or higher ($p = 0.003$). MRI-MRI exam pairs had more stable LI-RADS categorization compared to CT-CT (OR = 0.460, $p = 0.0018$).

Conclusions LI-RADS observations demonstrate increasing risk of progression to malignancy with increasing category ranging from 0% for LR-1 to 32% for LR-4 at 6 months. Inter-exam modality changes are associated with LI-RADS category changes.

Key Points
• *While the majority of LR-2 observations remain stable over long-term follow-up, LR-3 and especially LR-4 observations have a higher risk for category progression.*
• *Category transitions between sequential exams using different modalities (CT vs. MRI) may reflect modality differences rather than biological change. MRI, especially with the same type of contrast agent, may provide the most reproducible categorization, although this needs additional validation.*
• *In a clinical practice setting, in which radiologists refer to prior imaging and reports, there was no significant association between changes in radiologist and changes in LI-RADS categorization.*

✉ Claude B. Sirlin
csirlin@ucsd.edu

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

[2] School of Medicine, University of California San Diego, San Diego, CA, USA

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

[4] Division of Gastroenterology, Department of Medicine, University of California San Diego, San Diego, CA, USA

🖄 Springer

Introduction

Contrast-enhanced CT and MR imaging are essential in the clinical management of hepatocellular carcinoma (HCC) and are commonly used to diagnose this malignancy non-invasively [1]. Although several organizations had proposed imaging criteria for the diagnosis of HCC in at-risk patients [2], there was until recently little standardization regarding image interpretation and reporting for observations that do not meet those criteria. With the Liver Imaging Reporting and Data System (LI-RADS), the American College of Radiology attempts to standardize the categorization and reporting of the entire spectrum of imaging observations in patients at high-risk for HCC [3].

In LI-RADS, each possible lesion or region of concern is denoted an observation. Based on a combination of major and ancillary imaging features, each observation is assigned a category reflecting its relative likelihood of being HCC. LR-1 (definitely benign) observations include cysts and classic hemangiomas. LR-5 (definitely HCC) observations demonstrate arterial phase hyperenhancement in conjunction with other imaging features, such as capsule appearance and wash-out appearance, that together are diagnostic of HCC. The LR-5 criteria are intended to have 100% specificity for HCC and are generally consistent with those used by the American Association for the Study of the Liver [1] and the Organ Procurement and Transplantation Network [2, 4, 5]. LR-2 (probably benign), LR-3 (intermediate probability of malignancy), and LR-4 (probably HCC) are observations that are thought to have an intermediate and increasing likelihood of HCC. LR-M (probably or definitely malignant, not HCC specific) observations have features suggestive of malignancy but not specific for HCC.

The criteria for LI-RADS categorization were based in part on expert opinion and in part on evidence; however, literature at the time was limited by the lack of standardized terminology [6]. There has been an increasing body of evidence based on pathological and composite endpoints that suggests that higher LI-RADS categories correspond to an increasing probability of HCC [7–12].

However, there have been relatively few studies assessing the natural history of LI-RADS categories [13–16], and these studies have typically focused on assessing the risk of LR-4 progression to LR-5 or LR-M [15, 16]. Although Tanabe et al previously demonstrated that observations categorized as LR-2, LR-3, and LR-4 had different imaging outcomes, the

study was limited by its relatively small sample size. In addition, LI-RADS categories in these studies were assigned by performing dedicated research reads, and clinical radiology reports were not analyzed. The natural history of LR-1, LR-2, LR-3, and LR-4 observations, as reported in actual clinical practice, is still incompletely understood.

The purpose of this study is to better understand the natural history of these observations in the clinical setting by assessing the risk of progression of LR-1, LR-2, LR-3, and LR-4 observations. Since changes in radiologist or modality between exams may affect LI-RADS categorization and introduce longitudinal category transitions, an additional purpose was to assess the effects of inter-exam changes in radiologist or modality on LI-RADS categorization.

Methods

Study population

This is a retrospective analysis of consecutive clinical CT and MRI exams reported using standard LI-RADS version 2014 templates by faculty radiologists at our institution between January 2014 and September 2017. The study was Health Insurance Portability and Accountability Act compliant, and retrospective data collection and analysis were approved by our Institutional Review Board with a waiver of written informed consent. Standard LI-RADS v2014 CT and MRI templates used during the study period are found in the Supplementary Materials.

Data elements (study date, modality, radiologist, observation identifier, and observation category) were extracted automatically from reports for each reported observation using a custom Python script (version 2.7.10). All untreated LR-1, LR-2, LR-3, and LR-4 observations reported with at least one follow-up imaging exam reported with standardized LI-RADS templates were included in the study; all exams assessing included observations were analyzed after excluding exams that did not adhere to LI-RADS technical requirements. The first exam within the study period was considered the index exam, while the last was considered the final exam. No research reads were performed as part of the formal analysis and no prospectively reported categories were changed.

Imaging technique

Multiphase CT and MRI exams adhered to LI-RADS technical requirements

CT was performed with 64- and 320-detector row scanners, and MRI was performed at 1.5 T and 3 T. Pre-contrast, late hepatic arterial, portal venous, and delayed phases were acquired. CT examinations were performed by obtaining axial

images at 120 kVp and 200–750 mAs (adjusted for patient size), a table speed of 39.4 mm/rotation, and pitch of 0.987. A section thickness of 0.625 or less commonly 0.5 mm was used, and images were reconstructed with 2.5–3.75-mm axial slices and reformatted with 3–4-mm coronal and sagittal slices.

MRI included 3D fat-suppressed dynamic T1W with a flip angle of 15° and 4–6-mm section thickness with 50% overlap, single-shot T2W with coronal and axial acquisitions, in- and out-of-phase 3D T1W images, and diffusion imaging with low $(0–50 \text{ s/mm}^2)$ and intermediate-high $(300–1000 \text{ s/mm}^2)$ b values. MRI examinations were performed using gadobutrol, gadobenate dimeglumine, or gadoxetate disodium as the contrast agent. Hepatobiliary phase images were acquired only if gadoxetate disodium was used.

Additional technical details are provided in Supplementary Table 1 (CT) and Supplementary Table 2 (MRI).

Statistical analysis

Cohort, observation, modality, and radiologist characteristics were summarized descriptively. Continuous variables were reported as means and standard deviations, or as medians and inter-quartile ranges (IQR), as appropriate.

Two sets of analyses were done:

Set 1—Natural history. Category transitions between index and final exams were determined for each observation and summarized descriptively. Cumulative incidence curves for progression to a malignant LI-RADS category (LR-5 or LR-M) were computed for index LR-1, LR-2, LR-3, and LR-4 observations. Additionally, cumulative incidence curves for progression to LR-4 or higher (i.e., LR-4, LR-5, or LR-M) were computed for index LR-1, LR-2, and LR-3 observations. For these analyses, the time of conversion was estimated at the midpoint of the interval from the first reported instance of the outcome category (malignant LI-RADS category for the first set of analyses, LR-4 or higher for the second) and the nearest antecedent exam with the non-malignant category [17].

Using log-rank tests with a resampling extension to adjust for the variable number of observations per subject, cumulative incidences of progression to each category outcome were compared overall and, using Bonferroni's correction, pairwise. At each resampling iteration, one observation per patient was selected at random, and the cumulative incidence analysis with both overall and pairwise comparisons was performed. Test statistics were averaged over the iterations, and average log-rank test p values were computed based on the averaged statistics. For the analysis of progression to the malignant LI-RADS category p values $< 0.05/6$ were accepted as significant to ensure the family-wise 0.05 significance level.

For the analysis of progression to LR-4 or higher, p values $< 0.05/3$ were accepted as significant.

Set 2—Inter-exam transitions. For every sequential pair of exams for every observation, we calculated the inter-exam time interval as a continuous variable and recorded any inter-exam changes in LI-RADS category, changes in radiologist(s), and pair of modalities.

Modalities were classified as CT, MRI with an extracellular agent (ECA) or gadobenate dimeglumine, or MRI with gadoxetate disodium. The percentages of exams where there was no change in category, a change in category, an upgrade in category, and a downgrade in category, were summarized for each of the nine possible pairs of three modalities. Similarly, these percentages were summarized for sequential exams in which there was or was not a change in radiologist.

A multivariable mixed effect logistic regression analysis modeled the relationship between inter-exam changes in category and inter-exam directional changes in imaging modality (CT to CT which was default, CT to MRI, MRI to CT, or MRI to MRI) or radiologist. Patient-specific intercepts were fitted to adjust for within-patient dependence, and inter-exam time interval in years was included as a covariate in both models to adjust for time-dependent effects. Odds ratios (OR) were computed for model predictors.

Statistical analyses were performed with R version 3.3.3 (R Foundation for Statistical Computing). A significance level of 0.05 was used.

Results

Cohort, observation, modality, and radiologist characteristics

Our institution performed 2839 liver imaging exams (1031 CT, 1808 MRI) reported with LI-RADS v2014 during the study period. We identified all consecutive observations meeting eligibility criteria, excluding three exams that were performed non-contrast or using ferumoxytol. These comprised a total of 372 untreated LR-1, LR-2, LR-3, LR-4 observations with follow-up exams in 214 patients (149 male, 65 female; mean age 61.4 ± 9.9 years). The index exam was CT for 113 (30.4%) observations and MRI for 259 (69.6%) observations. Each observation was assessed on 2–8 exams (median = 2); 70 (18.8%) using CT only, 210 (56.5%) using MRI only, and 92 (24.7%) using both modalities depending on the time point. The median inter-exam follow-up interval was 124 days (IQR: 86–201 days), and the median total follow-up duration was 227 days (IQR 111–417 days).

These observations were assessed on a total of 715 exams (510 MRI, 205 CT). Of the 510 MRI exams, 388 (76.1%)

were performed with an extracellular agent (ECA), 122 (23.9%) were performed with gadoxetate disodium. Nineteen faculty radiologists reported clinical exams included in this study. All but three had year-long fellowship training in abdominal imaging, and 99.6% of exams (712/715) were reported by fellowship-trained abdominal radiologists. At the time of their first LI-RADS report included in this study, they ranged in post-residency experience from 12 months to over 30 years. Five radiologists reported 72% (517/715) of the exams, and 10 radiologists reported 94% (669/715) of the exams. Taking into account all index and follow-up exams, each observation was reported by 1–7 faculty radiologists (median = 2).

Natural history

Category transitions for the 372 index LR-1 to LR-4 observations between the index and final exams are summarized in Fig. 1 and illustrated in Fig. 2.

Of 10 index LR-1 observations, 5 (50%) remained LR-1 and 5 (50%) were LR-2 on final follow-up, and none progressed to LR-3 or higher. Of 43 index LR-2 observations, 5 (12%) decreased in category to LR-1, 30 (70%) remained LR-2, 4 (9%) progressed to LR-3, 2 (5%) progressed to LR-4, and 2 (5%) progressed to LR-5 ($n = 1$) or LR-M ($n = 1$). Of 186 index LR-3 observations, 2 (1%) decreased in category to LR-1, 27 (15%) decreased in category to LR-2, 107 (58%) remained LR-3, 33 (18%) progressed to LR-4, and 17 (9%) progressed to LR-5 ($n = 16$) or LR-M ($n = 1$). Of 133 LR-4 observations, none decreased in category to LR-1, 7 (5%) decreased in category to LR-2, 20 (15%) decreased in category to LR-3, 59 (44%) remained LR-4, and 47 (35%) progressed to LR-5 ($n = 44$) or LR-M ($n = 3$). No observations progressed to LR-5V.

Two LR-2 observations progressed to LR-5 or LR-M and were reviewed retrospectively by the first author (first-year radiology resident) and senior author (fellowship-trained abdominal radiologist with > 15 years of experience) in consensus. Both observations had mild-moderate T2 hyperintensity and restricted diffusion and, in retrospect, should have been
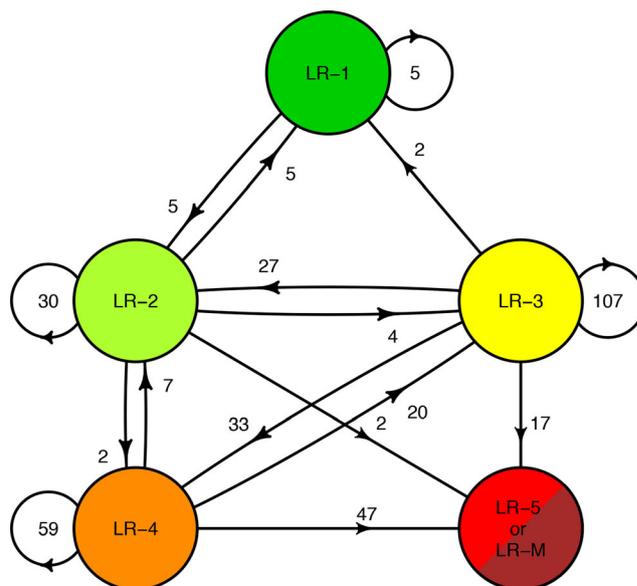

Fig. 2 Flow diagram illustrating the counts of each observation transition at the end of follow-up. Although the most of observations remained stable, observations exhibited significantly different rates of progression to higher LI-RADS categories based on their index category

categorized LR-3 based on ancillary features using v2014. Figure 3 illustrates the sequential category progression for one of these observations.

Overall, the cumulative incidence of progression to a malignant category differed significantly between LR-1, LR-2, LR-3, and LR-4 observations (Fig. 4, $p < 0.0001$). For LR-4, the cumulative incidence of progression to LR-5 or LR-M was 25%, 32%, 44%, and 46% at 3 months, 6 months, 1 year, and 2 years respectively. The corresponding rates were 3%, 7%, 11%, and 15% for LR-3 and 2%, 2%, 6%, and 6% for LR-2. In Bonferroni-corrected pairwise comparisons with an adjusted significance level of 0.0083, LR-4 had a significantly higher cumulative progression to a malignant category than LR-3 ($p < 0.0001$) or LR-2 ($p = 0.0006$) but not LR-1 ($p = 0.102$), although the small number of index LR-1s limited the power of this last comparison. Pairwise differences were not significant for LR-1 vs LR-2, LR-1 vs LR-3, or LR-2 vs LR-3 ($p = 0.629$, $p = 0.420$, and $p = 0.245$, respectively).

Fig. 1 LI-RADS category transitions summarizing index (rows) and final categories (columns) of 372 unique LR-1 to LR-4 observations are shown. Cells corresponding to observations that were stable in category are color-coded. Percentages are shown in parentheses for non-zero counts

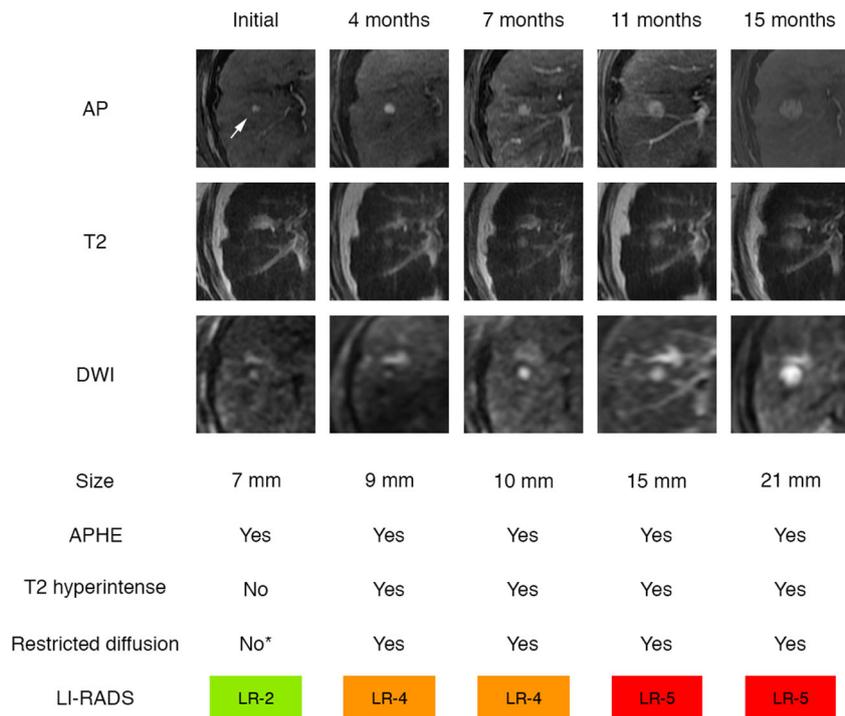|		Final category (median 7.5, up to 32 months later)				
		LR-1	LR-2	LR-3	LR-4	LR-5/-M
Index category	LR-1 N=10	5 (50%)	5 (50%)	0	0	0
	LR-2 N=43	5 (12%)	30 (70%)	4 (9%)	2 (5%)	2 (5%)
	LR-3 N=186	2 (1%)	27 (15%)	107 (58%)	33 (18%)	17 (9%)
	LR-4 N=133	0	7 (5%)	20 (15%)	59 (44%)	47 (35%)

Fig. 3 Category progression and imaging feature evolution as extracted from sequential radiology reports. Observation in hepatic segment 5/8 measured 7 mm, demonstrated arterial phase hyperenhancement (APHE, arrow) without definite washout or capsule appearance, and was categorized LR-2 on index exam. The observation subsequently grew, evolved in its imaging features, and progressed in LI-RADS category on follow-up exams at 4, 7, 11, and 15 months. In retrospect, the observation was hyperintense on high b value diffusion-weighted image (DWI) on index exam, indicating restricted diffusion. The observation should have been categorized LR-3, but this ancillary feature was unrecognized by the radiologist (indicated by an asterisk). AP: arterial phase images; T2: T2-weighted images. T2 hyperintense: mild-moderate T2 hyperintensity

Overall, the cumulative incidence of progression to LR-4 or higher differed significantly between LR-1, LR-2, and LR-3 observations (Fig. 5, $p = 0.041$). For LR-3, the cumulative incidence of progression to LR-4 or higher was 10%, 21%, 35%, and 39% at 3 months, 6 months, 1 year, and 2 years respectively. For LR-2, the corresponding rates were 5%, 7%, 12% and 12%, respectively. In Bonferroni-corrected pairwise comparisons with an adjusted significance level of 0.0167, LR-3 had a significantly higher cumulative progression to LR-4 or higher than LR-2 ($p = 0.0121$) but not LR-1 ($p = 0.148$). The pairwise difference between LR-1 vs LR-2 was not significant ($p = 0.475$).

Inter-exam transitions

The 372 observations were evaluated on a total of 724 sequential pairs of exams. In 96.4% (698/724) pairs, the two exams were performed at least 30 days apart. Each exam was performed either with CT, MRI with an ECA, or MRI with gadoxetate disodium. The percentages of exams where there was no change in category, a change in category, an upgrade in category, and a downgrade in category for each of the nine possible modality pairs are summarized in Table 1. Overall, the modality pairs with the most stable categorization were

MRI with the same type of contrast agent; gadoxetate disodium to gadoxetate disodium transitions had stable LR
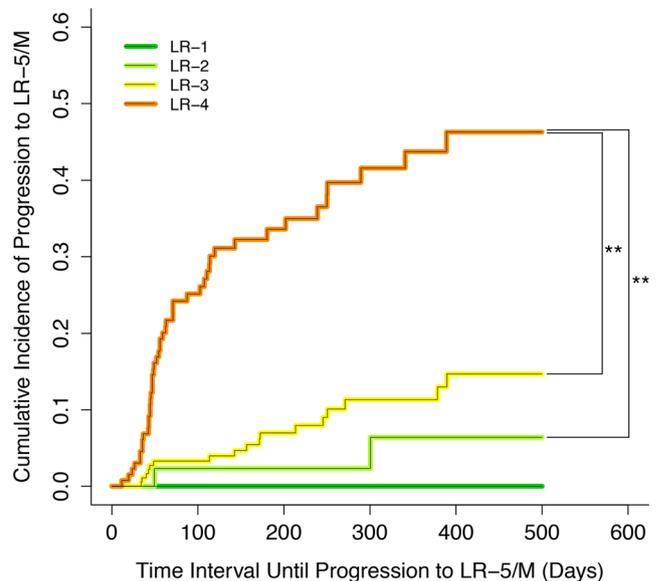

Fig. 4 Cumulative incidence curves for progression to LR-5 or LR-M shown for LR-1 (green), LR-2 (light green), LR-3 (yellow), and LR-4 (orange). Vertical bars indicate statistically significant pairwise comparisons assessed using Bonferroni-corrected log-rank tests with a resampling extension
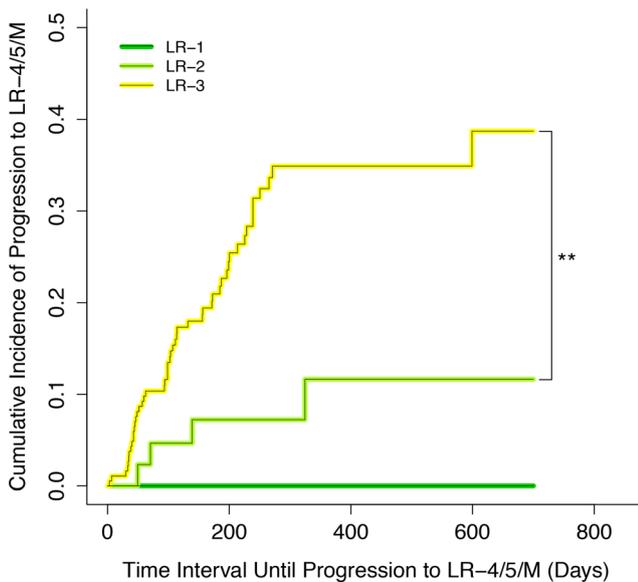
Fig. 5 Cumulative incidence curves for progression to LR-4 or higher shown for LR-1 (green), LR-2 (light green), and LR-3 (yellow). Vertical bars indicate the statistically significant pairwise comparison assessed using a Bonferroni-corrected log-rank test with a resampling extension

categorization in 80.4% of pairs, and ECA to ECA had stable categorization in 78.0% of pairs (Table 1).

The percentages of exams where there was no change in category, a change in category, an upgrade in category, and a downgrade in category, stratified by whether there was a change in radiologist, are summarized in Table 2. Categorization was stable in 74.6% pairs where there was no change in radiologist and in 69.1% pairs where there was a change (Table 2).

Multivariable mixed effect logistic regression for predictors of category change using follow-up interval, modality pair (CT to CT, CT to MRI, MRI to CT, MRI to MRI), and change in radiologist found that follow-up time was a

significant predictor of change (OR = 1.77/year, $p = 0.0294$); change in radiologist was not significantly associated with changes in LR category (OR = 1.18, $p = 0.503$). Compared to CT to CT (the default modality pair), MRI to MRI was a significant predictor of category stability (OR = 0.460, $p = 0.0018$); other pairs [CT to MRI (OR = 1.05, $p = 0.894$) and MRI to CT (OR = 0.895, $p = 0.755$)] were not significant predictors. Odds ratios and p values are summarized in Table 3.

Discussion

This study demonstrates that LR-1, LR-2, LR-3, and LR-4 observations have significantly different longitudinal outcomes. LR-1 observations have no risk of progression to malignant categories, whereas LR-2, LR-3, and LR-4 have an increasing likelihood of progression to malignant categories: in particular, LR-2 observations have a cumulative incidence of progression to LR-5 or LR-M of 2%, 2%, 6%, and 6% at 3 months, 6 months, 1 year, and 2 years, respectively; LR-3 observations have a cumulative incidence of progression to LR-5 or LR-M of 3%, 7%, 11%, and 15% at 3 months, 6 months, 1 year, and 2 years, respectively; and LR-4 observations have a cumulative incidence of progression to LR-5 or LR-M of 25%, 32%, 44%, and 46% at 3 months, 6 months, 1 year, and 2 years, respectively.

Because LR-4 observations have a high frequency of progression, optimal management is complex and probably warrants multidisciplinary discussion [18]. Prior studies estimating the likelihood of HCC and malignancy for various LI-RADS categories suggest that 64–87% of LR-4 observations are HCCs [8–12]. Depending on the clinical context and patient preferences, reasonable options may include close

Table 1 Pairs of exams separated by changes in LR and changes in modality (CT vs. MRI with extracellular agent vs. MRI with hepatobiliary agent). Percentages of exams where there was no change in LR category, a change in LR category, an upgrade in LR category, and a downgrade in LR category by whether there was a change in modality are shown. Median and IQR (inter-quartile range) of follow-up time as well as the proportion in each group that had a change in radiologist are also shown. *ECA* extracellular agent (including gadobenate dimeglumine), *Gx* gadoxetate disodium

Modality pair	Median follow-up time/days (IQR)	Proportion with change in radiologist (n)	Proportion with stable LR category (n)	Proportion with change in LR category (n)	Proportion with upgrade in LR category (n)	Proportion with downgrade in LR category (n)
CT to CT (n = 135)	104 (68–184.5)	83.7% (113/135)	60.7% (82/135)	39.3% (53/135)	24.4% (33/135)	14.8% (20/135)
CT to ECA (n = 56)	84 (34–122.5)	92.9% (52/56)	62.5% (35/56)	37.5% (21/56)	25.0% (14/56)	12.5% (7/56)
CT to Gx (n = 9)	90 (47–103)	100% (9/9)	44.4% (4/9)	55.6% (5/9)	22.2% (2/9)	33.3% (3/9)
ECA to CT (n = 45)	87 (54–120)	84.4% (38/45)	68.9% (31/45)	31.1% (14/45)	22.2% (10/45)	8.9% (4/45)
ECA to ECA (n = 304)	111 (85–174)	84.2% (256/304)	78.0% (237/304)	22.0% (67/304)	11.2% (34/304)	10.9% (33/304)
ECA to Gx (n = 43)	97 (94–220)	60.5% (26/43)	67.4% (29/43)	32.6% (14/43)	20.9% (9/43)	11.6% (5/43)
Gx to CT (n = 18)	75 (54–118.5)	100.0% (18/18)	50.0% (9/18)	50.0% (9/18)	22.2% (4/18)	27.8% (5/18)
Gx to ECA (n = 63)	157 (94–219)	81.0% (51/63)	63.5% (40/63)	36.5% (23/63)	28.6% (18/63)	7.9% (5/63)
Gx to Gx (n = 51)	193 (115.5–250)	84.3% (43/51)	80.4% (41/51)	19.6% (10/51)	11.8% (6/51)	7.8% (4/51)

Table 2 Pairs of exams separated by changes in LR category and changes in radiologist. Percentages of exams where there was no change in LR category, a change in LR category, an upgrade in LR category, and a downgrade in LR category by whether there was a change in radiologist are shown. Median and IQR (inter-quartile range) of follow-up times are also shown

	Median follow-up time/days (IQR)	Proportion with stable LR category (n)	Proportion with change in LR category (n)	Proportion with upgrade in LR category (n)	Proportion with downgrade in LR category (n)
No change in radiologist (n = 118)	93 (65.3–160)	74.6% (88/118)	25.4% (30/118)	19.5% (23/118)	5.9% (7/118)
Change in radiologist (n = 606)	112 (82.3–189.8)	69.1% (419/606)	30.8% (187/606)	17.8% (108/606)	13.0% (79/606)

imaging follow-up, alternative imaging, biopsy, or even treatment as presumptive HCC without biopsy confirmation. If imaging follow-up is pursued, our results suggest that follow-up intervals of 3 months and 6 months had a 25% and 32% risk of progression to a malignant category respectively.

In contrast, only 7% of LR-2 observations progressed to LR-4 or higher within 180 days, and the two LR-2 observations in our study that progressed to LR-5, on retrospective un-blinded review, had ancillary features favoring malignancy and should have been categorized LR-3. This suggests that follow-up every 6 months is reasonable, as the rate of progression during this interval is low. All LR-2 observations that progressed to LR-4 or higher did so within a year, suggesting that LR-2 observations that remain stable in category for at least a year are associated with a low risk of category progression; patients with such observations potentially could return to regular surveillance. There were only 10 LR-1 observations as these observations are infrequently reported in clinical practice.

The risks of progression to a malignant category for index LR-1, LR-2, LR-3, and LR-4 observations (0%, 5%, 9%, and 35%, respectively) are in keeping with other studies that assessed the longitudinal outcomes of LI-RADS categories [13–16]. Although two LR-2 observations in our study eventually progressed to LR-5, on retrospective un-blinded review, both had ancillary features favoring malignancy and should have been categorized LR-3. In prior studies, LR-2, LR-3, and

LR-4 had a 0%, 6–9%, and 31–38% rate of progression respectively to LR-5 or LR-M [13–16]. Our study helps to validate their findings in a larger and independent cohort. Moreover, unlike prior studies which categorized observations in the research setting by the consensus of two or more study radiologists, we extracted data elements directly from clinical radiology reports, a more realistic reflection of real-world clinical practice. Our study was also able to assess the association between changes in imaging modality and radiologists and the changes in LI-RADS categorization.

The directional change in modality was found to be significantly associated with changes in the LR category, even after adjusting for inter-exam time intervals and changes in radiologist. This may in part be related to inter-modality variability [19, 20]; however, a change in modality may also be the consequence of an observation being sub-optimally assessed on the initial exam. In this situation, the change in LI-RADS category may reflect a more accurate categorization. MRI to MRI modality pairs had the lowest proportion of category changes, and among MRI to MRI modality pairs, the modality pairs using the same MRI contrast agent had the lowest proportion of category changes. These results suggest that MRI, particularly with the same type of contrast agent, may provide the most reproducible categorization, although this needs to be validated in future studies.

We did not find an association between changes in radiologist with changes in LI-RADS category, despite the variability in LI-RADS scoring reported by prior reader reliability studies,

Table 3 Multivariable mixed effect logistic regression illustrating the relationship between directional changes in modality, changes in physician, and follow-up time interval in years. CT to CT and no change in physician were the default categories for the regression. *OR* odd ratio, *CI* confidence interval. OR for time interval is expressed as odds ratio per year of inter-exam interval

Predictor		OR (95% CI) relative to default predictor	p value
Modality pair	CT to CT transition (default)	–	–
	CT to MRI transition	1.055 (0.534–2.09)	0.877
	MRI to CT transition	0.898 (0.449–1.797)	0.761
	MRI to MRI transition	0.456 (0.281–0.741)	0.00153
Change in radiologist	No change in radiologist (default)	–	–
	Change in radiologist	1.122 (0.687–1.833)	0.645
Time interval (years)		1.769 (1.062–2.947)	0.0284

with one large international study finding an intra-class correlation coefficient of 0.67 for LI-RADS categorization [21–25]. The apparent lack of influence of change in radiologist on categorization may reflect anchoring bias towards prior categorizations in the clinical setting where radiologists refer to prior imaging and reports [26, 27].

Our study had several limitations. Because we analyzed only patients with follow-up imaging, observations that were treated or lost to follow-up could not be assessed, resulting in selection bias. Observation sizes and imaging features were not reported in a structured manner and thus not analyzed. In addition, although our institutional standard of care is to perform follow-up exams every 3–6 months, substantial variations in follow-up frequency and intervals depending on the patient's entire clinical context may confound direct comparison of progression risks. As this was a retrospective study, we could not enforce standardized follow-up intervals. No histological confirmation was performed, although prior studies have suggested that LR-5 and LR-M observations have a 98–100% and 92–100% probability of malignancy, respectively, and 95–100% and 25–77% probability of being HCC [7–12]. The study was performed at a single academic center, which may limit generalizability to populations in other geographical regions. As v2014 was the operational system during the study period, we were unable to assess longitudinal clinical outcomes with the recently released LI-RADS v2018. Thus, future studies should assess the outcomes of LI-RADS v2018 categories, ideally in larger and geographically diverse cohorts with standardized indications for modality changes and fixed follow-up intervals. Future studies could also assess the associations between multiple observations within the same patient and predictors of progression.

In summary, LI-RADS observations categorized in the clinical setting demonstrate the increasing risk of progression to malignancy with increasing category. While LR-1 observations have no risk of progression to malignancy and the majority of LR-2 observations remain stable, LR-3 and especially LR-4 observations have a higher potential for category progression. The optimal management of LR-4 observations is complex and may warrant multidisciplinary discussion. Category transitions between sequential exams performed using different modalities may in part reflect modality differences rather than biological change.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Cheng William Hong, MD MS.

Conflict of interest The authors of this manuscript declare relationships with the following companies:

Claude B Sirlin, MD, is a member of the external advisory board of AMRA Medical and Guerbet and speaker for GE Healthcare and Bayer (fees paid to University of California Regents); a consultant for Boehringer Ingelheim; receives research grants from Gilead, GE Healthcare, Siemens, GE MRI, Bayer AMRI, GE Digital, GE US, and ACR Innovation; gives education presentations for Medscape; and performs contracted research for ICON Medical Imaging/Enanta, Philips, Gilead, Shire, Virtualscopics, Intercept, and Synageva.

Statistics and biometry Tanya Wolfson, MA, and Anthony Gamst, PhD, kindly provided statistical advice for this manuscript. Both of these authors have significant statistical expertise.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology
• retrospective
• observational
• performed at one institution

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

References

1. Heimbach JK, Kulik LM, Finn RS et al (2018) AASLD guidelines for the treatment of hepatocellular carcinoma. Hepatology 67:358–380
2. Mitchell DG, Bruix J, Sherman M, Sirlin CB (2015) LI-RADS (liver imaging reporting and data system): summary, discussion, and consensus of the LI-RADS management working group and future directions. Hepatology 61:1056–1065
3. Kielar AZ, Elsayes KM, Chernyak V, Tang A, Sirlin CB (2018) LI-RADS version 2018: what is new and what does this mean to my radiology reports? Abdom Radiol (NY). https://doi.org/10.1007/s00261-018-1730-x
4. Elsayes KM, Kielar AZ, Agrons MM et al (2017) Liver imaging reporting and data system: an expert consensus statement. J Hepatocell Carcinoma 4:29–39
5. Wald C, Russo MW, Heimbach JK, Hussain HK, Pomfret EA, Bruix J (2013) New OPTN/UNOS policy for liver transplant allocation: standardization of liver imaging, diagnosis, classification, and reporting of hepatocellular carcinoma. Radiology 266:376–382
6. 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 286:29–48
7. Chen N, Motosugi U, Morisaka H et al (2016) Added value of a gadoxetic acid-enhanced hepatocyte-phase image to the LI-RADS system for diagnosing hepatocellular carcinoma. Magn Reson Med Sci 15:49–59
8. Choi SH, Byun JH, Kim SY et al (2016) Liver imaging reporting and data system v2014 with gadoxetate disodium-enhanced magnetic resonance imaging: validation of LI-RADS category 4 and 5 criteria. Invest Radiol 51:483–490
9. Cha DI, Jang KM, Kim SH, Kang TW, Song KD (2017) Liver imaging reporting and data system on CT and gadoxetic acid-enhanced MRI with diffusion-weighted imaging. Eur Radiol 27:4394–4405

10. Abd Alkhalik Basha M, Abd El Aziz El Sammak D, El Sammak AA (2017) Diagnostic efficacy of the Liver Imaging-Reporting and Data System (LI-RADS) with CT imaging in categorising small nodules (10–20 mm) detected in the cirrhotic liver at screening ultrasound. Clin Radiol 72:901.e1–901.e11

11. Kim YY, An C, Kim S, Kim MJ (2018) Diagnostic accuracy of prospective application of the Liver Imaging Reporting and Data System (LI-RADS) in gadoxetate-enhanced MRI. Eur Radiol 28: 2038–2046

12. Liu W, Qin J, Guo R et al (2018) Accuracy of the diagnostic evaluation of hepatocellular carcinoma with LI-RADS. Acta Radiol 59: 140–146

13. Tanabe M, Kanki A, Wolfson T et al (2016) Imaging outcomes of Liver Imaging Reporting and Data System version 2014 category 2, 3, and 4 observations detected at CT and MR imaging. Radiology 281:129–139

14. Choi JY, Cho HC, Sun M, Kim HC, Sirlin CB (2013) Indeterminate observations (Liver Imaging Reporting and Data System category 3) on MRI in the cirrhotic liver: fate and clinical implications. AJR Am J Roentgenol 201:993–1001

15. Burke LMB, Sofue K, Alagiyawanna M et al (2016) Natural history of liver imaging reporting and data system category 4 nodules in MRI. Abdom Radiol (NY) 41:1758–1766

16. Sofue K, Burke LMB, Nilmini V et al (2017) Liver imaging reporting and data system category 4 observations in MRI: risk factors predicting upgrade to category 5. J Magn Reson Imaging 46:783–792

17. Law CG, Brookmeyer R (1992) Effects of mid-point imputation on the analysis of doubly censored data. Stat Med 11:1569–1578

18. Mitchell DG, Bashir MR, Sirlin CB (2017) Management implications and outcomes of LI-RADS-2, -3, -4, and -M category observations. Abdom Radiol (NY) 43:143–148

19. Zhang YD, Zhu FP, Xu X et al (2016) Liver Imaging Reporting and Data System:: substantial discordance between CT and MR for imaging classification of hepatic nodules. Acad Radiol 23:344–352

20. Chernyak V, Flusberg M, Law A, Kobi M, Paroder V, Rozenblit AM (2017) Liver imaging reporting and data system: discordance between computed tomography and gadoxetate-enhanced magnetic resonance imaging for detection of hepatocellular carcinoma major features. J Comput Assist Tomogr 42(1). https://doi.org/10.1097/RCT.0000000000000642

21. Fowler KJ, Tang A, Santillan C et al (2018) Interreader reliability of LI-RADS version 2014 algorithm and imaging features for diagnosis of hepatocellular carcinoma: a large international multireader study. Radiology 286:173–185

22. Davenport MS, Khalatbari S, Liu PS et al (2014) Repeatability of diagnostic features and scoring systems for hepatocellular carcinoma by using MR imaging. Radiology 272:132–142

23. Barth BK, Donati OF, Fischer MA et al (2016) Reliability, validity, and reader acceptance of LI-RADS—an in-depth analysis. Acad Radiol 23:1145–1153

24. Ehman EC, Behr SC, Umetsu SE et al (2016) Rate of observation and inter-observer agreement for LI-RADS major features at CT and MRI in 184 pathology proven hepatocellular carcinomas. Abdom Radiol (NY) 41:963–969

25. Zhang YD, Zhu FP, Xu X et al (2016) Classifying CT/MR findings in patients with suspicion of hepatocellular carcinoma: comparison of liver imaging reporting and data system and criteria-free Likert scale reporting models. J Magn Reson Imaging 43:373–383

26. Lee CS, Nagy PG, Weaver SJ, Newman-Toker DE (2013) Cognitive and system factors contributing to diagnostic errors in radiology. AJR Am J Roentgenol 201:611–617

27. Bruno MA, Walker EA, Abujudeh HH (2015) Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction. Radiographics 35:1668–1676