Diagnostic test accuracy systematic reviews: summarizing the evidence of diagnostic approaches in cancer-related imaging

Haben Tesfu Dawit

Thesis submitted to the University of Ottawa in partial fulfillment of the requirements for the Master of Science in Epidemiology

School of Epidemiology and Public Health
Faculty of Medicine
University of Ottawa

Preface

The thesis project was written by Haben Dawit under the supervision of Dr. Matthew McInnes.

Chapter 2 was published in the Journal of Magnetic Resonance Imaging in August 2022 [Frank RA, Dawit H, Bossuyt PMM, Leeflang M, Flood TA, Breau RH, McInnes MDF, Schieda N. Diagnostic Accuracy of MRI for Solid Renal Masses: A Systematic Review and Meta-analysis. J Magn Reson Imaging. Apr 2023;57(4):1172-1184. doi:10.1002/jmri.28397]. No ethics board approval was required for this study. I assisted in the design of the protocol, and led all aspects of data collection, data analysis and the writing of the manuscript. Robert Frank led the protocol design and assisted in the data collection, data analysis and the writing of the manuscript. All other authors contributed to the protocol design and the manuscript edits. Drs. McInnes and Schieda (corresponding author) provided guidance and supervision during all stages of the research project.

Chapter 3 was for publication to Radiology on June 15th, 2023.. The authors of Chapter 3 are Dawit H, Lam E, McInnes MDF, van der Pol CB, Bashir MR, Salameh JP, Levis B, Chernyak V, Sirlin CB, Choi SH, Kim SY, Fraum TJ, Tang A, Jiang H, Song B, Wang J, Wilson SR, Kwon H, Kierans AS, Joo I, Ronot M, Song JS, Podgórska J, Rosiak G, Kang Z, Allen BC and Costa AF. IRB approval was received from the Ottawa Hospital. I am the lead investigator of the project. I contributed to the design of the protocol and lead all aspects of data collection and data analysis with the supervision of Eric Lam (OHRI Methodologist). Dr. Andreu Costa drafted the original manuscript. Quality assessment was previously conducted by Dr. Christian van der Pol and Jean-Paul Salameh as part of the ongoing research conducted by the LI-RADS committee. The manuscript was drafted by Dr. Costa and I. All other authors assisted in the protocol design and manuscript edits. Drs. Bashir, van der Pol, McInnes and Costa (corresponding author) provided guidance during all stages of the research project.

Chapter 4 is an original, unpublished work by Haben Dawit. Drs. McInnes, Hutton and Breau provided manuscript edits.

Abstract

Systematic reviews on diagnostic test accuracy studies provide an overview of the current literature in a systematic and transparent manner. They are the highest level of evidence in clinical research, therefore they are critical to decision-making in the healthcare setting. Cancer is the primary source of mortality in Canada, however early detection of tumors can improve the survival rates and long-term health outcomes of patients. The primary method of cancer diagnosis is histopathological assessment, however, its use remains controversial. It is an invasive procedure and requires resources and clinical expertise not readily available. Non-invasive clinical imaging has been studied as a clinically desirable method for cancer diagnosis, however its diagnostic accuracy has yet to be established in the medical setting. With the increased role of DTA research in medicine, the current literature needs to be summarized in an effective way to properly educate and influence clinical decision-making. The objective of this thesis is to address the current evidence gaps in DTA research by conducting several systematic reviews to evaluate the accuracy of diagnostic methods in cancer-related imaging. The last chapter of the thesis will provide a critical reflection on the current landscape of DTA studies in cancer-related imaging, based on the findings of the reviews in the thesis.

Acknowledgements

My thesis could not have been completed without the invaluable guidance and support from those around me including my supervisor, colleagues and my family.

I would like to thank my supervisor, Dr. Matthew McInnes, without whom my research would not have been possible. You provided the information and resources needed during all stages of my degree. Furthermore, not only did you develop my knowledge and skills in clinical epidemiology, you went above and beyond to give me numerous opportunities to reach my fullest potential as a researcher. I am truly grateful for the dedication you showed towards my academic, professional and personal development.

I would like to thank my TAC members, Drs. Rodney Breau and Brian Hutton. You sacrificed your time and energy to provide me guidance and support throughout my thesis project.

I would like to thank my fellow authors on all my projects. You provided me with great insight into epidemiological methods and standardized practices in the clinical environment. I would like to extend my gratitude to Dr. Robert Frank, Dr. Andreu Costa and Dr. Nicola Schieda for their continued efforts throughout my projects.

I would like to thank my colleagues, especially Eric Lam, Nayaar Islam, Jean-Paul Salameh and Sanam Ebrahimzadeh. You guided me through many challenges and always made yourselves available when I required assistance. Your continued support was instrumental to the successful completion of my projects.

I would like to thank the University of Ottawa, the Ottawa Hospital Research Institute (OHRI) and the province of Ontario for providing financial support throughout my master's degree. These scholarships include the University of Ottawa's Admissions Scholarship, the Graduate Studentship for advanced training (OHRI) and the Queen Elizabeth II Graduate Scholarship in Science and Technology (QEII-GSST) Program.

Finally, I would like to thank my family for their endless support. Their sacrifices gave me the opportunity and confidence to pursue my academic goals. I would not be in the position I am today without them.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1. General Introduction

Diagnostic test accuracy (DTA) studies evaluate the performance of diagnostic tests in the medical field (1). These studies are based on the ability of a test to differentiate the presence or absence of a disease, compared to the current "gold" standard (2, 3). The results of DTA studies are summarized in DTA systematic reviews.

Systematic reviews are the highest level of evidence in clinical research (4). They provide an overview of the current literature, in a systematic and transparent manner (1). All applicable studies are assessed for quality, and if appropriate, the results are summarized using a meta-analysis (1, 3). The results of reviews are commonly presented using sensitivity and specificity, as well as other summary measures including forest plots, and summary receiver operating characteristics (sROC) curves (1). Sensitivity is defined as the probability of getting a positive result given a subject has the disease (5). Specificity is defined as the probability of obtaining a negative result given a subject does not have the disease (5). Both measures are strong indicators' of the clinical effectiveness of a test, as they are not influenced by the disease prevalence (1, 3, 5).

The goal of systematic reviews is to inform and support the delivery for evidence-based practice. The outcomes of these reviews play a role in the policies adopted in healthcare settings. Therefore, it is critical that reviews are conducted on current research to advance the knowledge and practices of medicine.

Cancer is the primary cause of death in Canada (6). In 2021, it was estimated that 229 200 people will be diagnosed with cancer, and 84 600 people will die from cancer (6). The increase in cancer diagnosis and deaths places a greater demand for diagnostic tests. The gold standard for diagnostic tests in cancer is pathological assessment (ex. Biopsy), however it is not clinically favorable (7-10). Procedures are highly invasive and requires resources and clinically expertise not universally available (7-10). The use of non-invasive imaging is clinically preferred for diagnosis, however studies evaluating the diagnostic accuracy of computed tomography (CT), magnetic resonance imaging (MRI) and other imaging modalities have yet to establish their accuracy in diagnosis. With the increased role of DTA research in medicine (11, 12), it is

essential that the effectiveness of current and novel diagnostic approaches are rigorously evaluated, to properly inform decision-making in the medical field.

Objective

To address the current evidence gaps in DTA research, several systematic reviews will be conducted to evaluate the accuracy of diagnostic methods in cancer-related imaging. The diagnostic accuracy systematic reviews will adhere to the methodologic guidance from the Cochrane DTA methods group (13) and the and is reported according to the preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data (PRISMA-IPD) and PRISMA for Diagnostic Test Accuracy (DTA) guidelines (14, 15).

Chapter descriptions

Chapter 2. Diagnostic accuracy of MRI for solid renal masses: A systematic review and meta-analysis.

The purpose of this systematic review is to summarize the available evidence on the diagnostic accuracy of bi-parametric MRI (bpMRI) and multi-parametric MRI (mpMRI) for diagnosis of malignant (versus benign) solid renal tumors. Secondary objectives are the evaluation of the accuracy for diagnosing aggressive (versus indolent) tumors and clear cell renal cell carcinoma (ccRCC) (versus any other histologic diagnosis) among solid renal masses, and to identify sources for variability in accuracy.

Chapter 3. Association of Liver Imaging Reporting and Data System (LI-RADS) CT/MRI Ancillary Features (AFs) with hepatocellular carcinoma (HCC) and Malignancy: Systematic Review and Individual Participant Data (IPD) Meta-Analysis

The purpose of this IPD meta-analysis is to evaluate the association of AFs with HCC and malignancy in a model that includes LI-RADS major features. The secondary objective is to assess the proportion of observations that exhibited each AF.

Chapter 4. Reflecting on the current landscape of diagnostic test accuracy studies in cancer-related imaging

This section summarizes the findings of the previous articles and evaluates their clinical impact. Furthermore, it evaluates pitfalls in current DTA research and how they could be addressed to improve the quality of imaging research.

References

1. Leeflang MM. Systematic reviews and meta-analyses of diagnostic test accuracy. Clin Microbiol Infect. 2014;20(2):105-13.
2. Sitch AJ, Dekkers OM, Scholefield BR, Takwoingi Y. Introduction to diagnostic test accuracy studies. Eur J Endocrinol. 2021;184(2):E5-E9.
3. Mander GTW, Munn Z. Understanding diagnostic test accuracy studies and systematic reviews: A primer for medical radiation technologists. J Med Imaging Radiat Sci. 2021;52(2):286-94.
4. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. Plast Reconstr Surg. 2011;128(1):305-10.
5. Šimundić AM. Measures of Diagnostic Accuracy: Basic Definitions. EJIFCC. 2009;19(4):203-11.
6. in CCSAC, collaboration with the Canadian Cancer Society S, Canada CatPHAo. Canadian Cancer Statistics 2021. Toronto,ON: C anadian Cancer Society; 2021.
7. Patel D, Ghali F, Patel S, Miller N, Bradshaw A, Meagher M, et al. Utilization of renal mass biopsy in patients with localized renal cell carcinoma: A population-based study utilizing the National Cancer Database. Journal of Clinical Oncology. 2020;38(6_suppl):662-.
8. Richard PO, Martin L, Lavallée LT, Violette PD, Komisarenko M, Evans AJ, et al. Identifying the use and barriers to the adoption of renal tumour biopsy in the management of small renal masses. Can Urol Assoc J. 2018;12(8):260-6.
9. Menogue SR, O'Brien BA, Brown AL, Cohen RJ. Percutaneous core biopsy of small renal mass lesions: a diagnostic tool to better stratify patients for surgical intervention. BJU international. 2013;111(4 Pt B):E146-51.
10. Patel HD, Johnson MH, Pierorazio PM, Sozio SM, Sharma R, Iyoha E, et al. Diagnostic Accuracy and Risks of Biopsy in the Diagnosis of a Renal Mass Suspicious for Localized Renal Cell Carcinoma: Systematic Review of the Literature. J Urol. 2016;195(5):1340-7.
11. Singh H, Graber ML. Improving Diagnosis in Health Care--The Next Imperative for Patient Safety. N Engl J Med. 2015;373(26):2493-5.
12. Care CoDEiH, Services BoHC, Medicine Io, The National Academies of Sciences Eg, and Medicine. Improving Diagnosis in Health Care. 2015.

13. Group CDM. Handbook for DTA Reviews 2021 [Available from: https://methods.cochrane.org/sdt/handbook-dta-reviews.
14. Cohen JF, Deeks JJ, Hooft L, Salameh JP, Korevaar DA, Gatsonis C, et al. Preferred reporting items for journal and conference abstracts of systematic reviews and meta-analyses of diagnostic test accuracy studies (PRISMA-DTA for Abstracts): checklist, explanation, and elaboration. Bmj. 2021;372:n265.
15. Salameh JP, Bossuyt PM, McGrath TA, Thombs BD, Hyde CJ, Macaskill P, et al. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. Bmj. 2020;370:m2632.

Chapter 2. Diagnostic accuracy of MRI for solid renal masses: A systematic review and meta-analysis.

List of authors

1. *Haben Dawit BSc. School of Epidemiology, Public Health and Preventative Medicine, Faculty of Medicine, University of Ottawa; Clinical Epidemiology Program, Ottawa Hospital Research Institute.
2. *Robert A. Frank MD. Department of Radiology, University of Ottawa
3. Patrick MM Bossuyt PhD. Amsterdam University Medical Centers, University of Amsterdam, Epidemiology and Data Science, Amsterdam, the Netherlands.
4. Mariska Leeflang PhD. Amsterdam University Medical Centers, University of Amsterdam, Epidemiology and Data Science, Amsterdam, the Netherlands.
5. Trevor A Flood MD. Department of Anatomical Pathology, University of Ottawa
6. Rodney H, Breau MD. University of Ottawa Department of Surgery. Clinical Epidemiology Program, Ottawa Hospital Research Institute.
7. Matthew DF McInnes MD. Department of Radiology, University of Ottawa; Clinical Epidemiology Program, Ottawa Hospital Research Institute.
8. Nicola Schieda MD. Department of Radiology, University of Ottawa; Clinical Epidemiology Program, Ottawa Hospital Research Institute.

*RAF and HD contributed equally and are joint first authors

Preface

Manuscript Status

Chapter 2 was published in the Journal of Magnetic Resonance Imaging in August 2022 [Frank RA, Dawit H, Bossuyt PMM, Leeflang M, Flood TA, Breau RH, McInnes MDF, Schieda N. Diagnostic Accuracy of MRI for Solid Renal Masses: A Systematic Review and Meta-analysis. J Magn Reson Imaging. Apr 2023;57(4):1172-1184. doi:10.1002/jmri.28397].

Objective

The purpose of this chapter is to synthesize the available evidence on the diagnostic accuracy of bp-MRI and mp-MRI for diagnosing solid renal masses.

Contributions of authors

- Conception of the study protocol: Frank RA, Dawit H, McInnes MDF Schieda N
- Study selection and data extraction: Dawit H, Frank RA
- Data validation: Dawit H, Frank RA
- Data analysis and interpretation: Dawit H
- Drafting of the manuscript: Dawit H, Frank RA
- Revision of the manuscript: Frank RA, Bossuyt PMM, Leeflang M, Flood TA, Breau RH, McInnes MDF, Schieda N
- Research guidance and support: McInnes MDF, Schieda N

Protocol registration

The study protocol was registered before initiation with PROSPERO (CRD42022303844).

Ethics approval

No ethics board approval was required for this study.

Appendices

Supplementary material is available at the end of this thesis in appendix 2.1-2.10 Please refer to list of appendices on page VIII for specific page numbers.

Funding

Dawit H was supported by the University of Ottawa's Admissions Scholarship and the Graduate Studentship for advanced training (OHRI)

Acknowledgements

We would like to thank Risa Shorr and Josée Skuce for assisting with the literature search and helping us obtain full-text articles. In addition, we would like to acknowledge the work of Jean-Paul Salameh, Nayaar Islam and Sanam Ebrahimzadeh for their continued support throughout the project.

Abstract

Background: Biparametric(bp)-MRI and multiparametric(mp)-MRI may improve the diagnostic accuracy of renal mass histology.

Purpose: To evaluate the available evidence on the diagnostic accuracy of bp-MRI and mp-MRI for solid renal masses in differentiating malignant from benign, aggressive from indolent, and clear cell renal cell carcinoma(ccRCC) from other histology.

Study Type: Systematic review

Population: MEDLINE, EMBASE, and CENTRAL up to January 11th,2022 were searched.

Field Strength/Sequence: NA

Assessment: Eligible studies evaluated the accuracy of MRI (with at least two sequences: T2, T1, dynamic contrast and diffusion-weighted imaging) for diagnosis of solid renal masses in adult patients, using histology as reference standard. Risk of bias and applicability were assessed using QUADAS-2.

Statistical Tests: Meta-analysis using a bivariate logitnormal random effects model.

Results: We included 10 studies (1239 masses from approximately 1200 patients). The risk of bias was high in three studies, unclear in five studies and low in two studies. The diagnostic accuracy of malignant (vs benign) masses was assessed in five studies (64%[179/281] malignant). The summary estimate of sensitivity was 95%(95% confidence interval(CI):77%-99%), and specificity was 63%(95%CI:46%-77%). No study assessed aggressive (vs indolent) masses. The diagnostic accuracy of ccRCC (vs other subtypes) was evaluated in six studies (47%[455/971]ccRCC): the summary estimate of sensitivity was 85%(95%CI:77%-90%) and specificity was 77%(95%CI:73%-81%).

Data Conclusion: Our study reveals deficits in the available evidence on MRI for diagnosis of renal mass histology. The number of studies was limited, at unclear/high risk of bias, with heterogeneous definitions of solid masses, imaging techniques, diagnostic criteria, and outcome measures.

Keywords

Diagnostic accuracy; systematic review; multi-parametric MRI; bi-parametric MRI; renal mass histology; clear cell renal cell carcinoma

Introduction

Renal cell carcinoma (RCC) is the most common kidney cancer and is frequently an incidental imaging finding (1-3). The increasing detection of renal tumours in imaging studies (4, 5) necessitated an increase in the number of surgeries and renal tumour ablations performed (6). However, renal cancer-specific mortality has remained relatively stable over the same time period (7). Therefore, active surveillance has been proposed as an alternative management option for incidentally discovered small renal masses in selected patients and is growing in popularity in clinical practice (1, 8-10).

Decision-making for treatment of renal mases is individualized and should ideally be patient and tumor-specific (11, 12). Currently, initial management decisions regarding active surveillance versus extirpative treatment generally do not consider the histological diagnosis of a mass, but instead rely primarily on patient age, comorbidities, and mass size ± growth rate, because histologic confirmation requires invasive needle biopsy (1). Among solid renal masses, a diagnosis of malignancy, and particularly aggressive histology among malignant tumors, may influence management decisions. Up to 20% of small solid renal masses are benign in large surgical series (13, 14). Even when malignant, many small renal tumors are clinically indolent with low rates of disease progression or metastasis, and the proportion of indolent RCCs is higher in smaller masses (15, 16). Predicting aggressive behavior can be challenging; however, adverse pathological features (e.g., higher grade, sarcomatoid differentiation, necrosis) are important for indicating aggressive disease (17).

Among the most common subtypes of RCC, clear cell RCC (ccRCC) is typically considered the most aggressive, with the highest risk of disease progression and metastasis (18). Therefore, pre-treatment knowledge of malignant histology, and particularly aggressive histology, could help inform the most appropriate patient management.

Pre-treatment histological diagnosis of renal masses can be achieved through renal mass biopsy (19). Biopsy generally has a low complication rate and high diagnostic accuracy, in adequate samples, for benign versus malignant disease and high pathological concordance to mass subtype after nephrectomy (20). However, biopsy is still invasive, requires technical expertise that is not universally available, adds another step to patient management, and has a

non-diagnostic rate of up to 14%, which contributes to underutilization in clinical practice (21-24). The use of non-invasive imaging to diagnose renal mass histology is therefore clinically desirable. Yet, to date, studies evaluating the diagnostic accuracy of both CT and MRI to diagnose solid renal mass subtypes are limited.

There is growing interest in the use of MRI, in particular MRI which combines two (bi-parametric MRI [bpMRI]) or more (multi-parametric [mpMRI]) sequences for diagnosis of renal mass subtypes (25), given the wealth of imaging data afforded by these imaging techniques (26).

Although previous studies have evaluated the morphological differences between malignant and benign renal masses, the accuracy of MRI for differentiating benign from malignant solid renal masses is inconsistent (27, 28).

The purpose of this systematic review was to summarize the available evidence on the diagnostic accuracy of bpMRI and mpMRI for diagnosis of malignant (versus benign) solid renal tumors. Secondary objectives were to evaluate the accuracy for diagnosing aggressive (versus indolent) tumors and clear cell RCC (versus any other histologic diagnosis) among solid renal masses, and to identify sources for variability in accuracy.

Materials and methods

This diagnostic accuracy systematic review adhered to the methodologic guidance from the Cochrane DTA methods group (29) and is reported according to the PRISMA-DTA guidelines (30, 31). Institutional ethics approval is not required for this study type at our institution. The study protocol was registered before initiation with PROSPERO (CRD42022303844).

Study Question

The following 'Patient, Index Test, Target Condition' (PIT) question was constructed.
Patients: Adult patients with solid renal masses
Index Test: MRI of the kidneys, with at least 2 clinically routine sequences (T2W, T1-chemical shift, dynamic contrast enhancement and/or diffusion weighted imaging).

Target Condition: Primary = malignant tumor (vs. benign); Secondary = aggressive tumor (vs. indolent) and ccRCC (vs any other solid mass).

Protocol Deviation

An earlier version of the study protocol planned for the assessment of diagnostic accuracy of mpMRI (use of 3 or more sequences) only; however, after initial search, it was determined that the number of studies with the appropriate index test (and meeting all other inclusion criteria) was minimal (6 total; 3 for the primary outcome and for the secondary outcome ccRCC vs. other). The study protocol was subsequently modified to include bi-parametric MRI as an index test.

Search

The search included MEDLINE, EMBASE, and Cochrane Central Register of Controlled Trials (CENTRAL), with the assistance of an experienced hospital librarian (search details are shown in Appendix 2.1). There was no date or language restriction. All original research except for case reports were included in the search; non-original research (reviews, commentary, letter to the editor) were excluded.

Initial screening of search results based on title and abstract was performed independently by two reviewers (RAF [4th year radiology resident] and HD [1st year graduate student]). Any study deemed potentially eligible by either reviewer was independently evaluated by the same 2 reviewers and the study was included if deemed eligible by both. Discrepancies were resolved by consensus discussion; persistent disagreements were resolved by discussion with a third reviewer (MM or NS [Staff radiologists, each with 10+ years of meta-research and abdominal imaging experience]. The reference lists from all included studies were checked for additional eligible citations. If studies contained overlapping study groups, the study with the largest sample size was used.

Inclusion Criteria

To be included in the review, a study was expected to meet the following criteria: the study group included adult patients with a solid renal mass (any definition of "solid" was acceptable), MRI of the kidneys was performed with at least two clinically routine sequences (T2

weighted, T1-chemical shift [T1 in phase and out of phase], dynamic contrast-enhanced (DCE) and diffusion-weighted imaging (DWI)), and the reference standard was histological assessment.

The final MRI interpretation could be either purely subjective (e.g. radiologist impression) or contain quantitative elements [e.g., region of interest (ROI) measurements, Signal Intensity ratios, DCE curves]. To be included, DCE sequences required pre-contrast, corticomedullary, and nephrographic phases. Studies that evaluated entirely quantitative diagnostic criteria (including ROI measurements) and non-human interpretation models (e.g., texture analysis, radiomics, multivariable prediction models, or similar) were excluded due to a lack of formal governing body administrative approval for routine clinical use. Studies that did not contain sufficient data to calculate a 2 x 2 contingency table were excluded due to concerns over reproducibility.

Index Test Interpretation

The definition of a 'positive' result was based on the criteria applied in each included primary study.

Reference Standard

For the primary target condition (malignant vs. benign), the diagnosis was considered 'positive' if the reported pathologic diagnosis was classified as a 'malignant tumor' as defined by the morphology code in the 2016 World Health Organization classification of renal tumors (Appendix 2.2) (32) and was considered 'negative' if any other pathologic diagnosis was reported. If the classification remained indeterminate based on the WHO classification, the criteria defined in a prior study by Bhindi et al. (17) was used as secondary classification tool (Appendix S2.3).

For the secondary target condition 'aggressive vs. indolent', the diagnosis was considered 'positive' if the reported pathologic diagnosis was an 'aggressive tumor' as defined in appendix 2.3 (17), and was considered 'negative' if any other pathologic diagnosis was reported.

For the secondary target condition of 'clear cell RCC vs. other', the diagnosis was considered 'positive' if the reported pathologic diagnosis was 'clear cell renal cell carcinoma' of any grade and was considered 'negative' if any other pathologic diagnosis was reported (including other RCC subtypes, any other malignancies, and any benign diagnosis).

When sufficient histological details were available, tumors were characterized according to the 2016 WHO classification system (32) and in the case of a discrepancy with the reported pathologic result in a primary study, the retrospectively determined 2016 WHO classification was used.

Data Extraction

After a pilot of the data extraction form on one study by two investigators (RAF and HD), the following data were extracted from included studies in duplicate: first author, year of publication, study details (data collection [before the index test and reference standard were applied vs after], patient enrolment [consecutive vs. convenience vs. random], study design [randomized trial vs. cohort vs. case-control], funding source), number of patients, number of masses, patient age (mean, range), definition of 'solid' mass, technical parameters of the index test (magnet strength, sequences performed, b-value of DWI if performed, number and timing of post-contrast phases if DCE performed), imaging features/criteria used to define a 'positive' diagnosis, reference standard details (biopsy vs partial nephrectomy vs. total nephrectomy), final pathologic/histologic diagnosis.

The number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) was extracted for construction of 2x2 contingency tables. For each outcome (target condition definitions as per above), TP was defined as: 'positive' by MRI criteria and 'positive' on final pathology, TN was defined as: 'negative' by MRI criteria and 'negative' on final pathology, FP was defined as 'positive' by MRI criteria and 'negative' on final pathology, and FN was defined as 'negative' by MRI criteria and 'positive' on final pathology.

Risk Of Bias And Applicability Assessment

Risk of bias was assessed using a customized version of the QUADAS-2 tool based on the four domains of study selection, index test, reference standard, and flow and timing (See

appendix 2.4 for the full QUADAS-2 criteria tailored for this study question). Concerns regarding applicability were also assessed using the QUADAS-2 tool (33). The customized tool was applied to all studies in duplicate by 2 investigators (RAF and HD) with disagreements resolved by consensus and discussion with a third investigator (MM or NS) when necessary. Selective reporting (e.g. non-reporting) was classified as high risk of bias, if the element not reported was critical to the inclusion of the study. The risk of publication bias was not assessed, as per contemporary guidance (34).

Data Analysis

The per-lesion 2x2 data were summarized in forest plots of sensitivity and specificity with 95% confidence intervals for each study. Summary receiver operating characteristic (SROC) plots with 95% confidence intervals and 95% prediction regions were constructed for each outcome using MetaDTA (35, 36). Summary estimates of the sensitivity and specificity of bpMRI or mpMRI for each target condition (malignancy, 'aggressive' tumor, and clear cell RCC) were obtained with a bivariate binomial random effects model, using a generalized linear mixed model approach (37, 38). The summary estimate of sensitivity and specificity represent the means of the underlying bivariate logitnormal distribution.

If accuracy results were presented for multiple readers in a single primary study, they were averaged across the readers and inter-observer variability was extracted from these studies (39). If multiple thresholds or criteria for positivity were reported in a single primary study, accuracy estimates were extracted separately and modeled using a linear mixed effects model for modeling multiple threshold data, as described in Schneider et al. (40-42).

All studies that used the clear cell likelihood score (ccLS) (a 5-point scoring system that incorporates mpMRI imaging features in diagnosing ccRCC) were defined as ccRCC "positive" if they had score of 4 or 5 (18, 43), while other masses (ccLS 1-3) were classified as "negative" (18, 43). The classification of ccLS scores as either ccRCC positive or negative was decided after data extraction, but in accordance with guidance from the authors who derived the system (44). If sufficient studies were available, the 2x2 data for individual thresholds would be modeled using the analysis for multiple thresholds described above. The diagnostic accuracy at each threshold score was summarized using an SROC plot.

Analyses were planned to assess sources of variability in the data. The diagnostic accuracy of bpMRI and mpMRI individually would have been assessed by stratifying studies according to the number and combination of sequences used in diagnosis. Stage cT1a (≤4 cm) masses were to be grouped together to determine the effect of renal mass size on diagnostic accuracy. Subgroups were planned according to differences in diagnostic criteria applied and parameters used in the MRI protocol (exact approach to stratification was determined based on heterogeneity of included primary studies). Lastly a sensitivity analysis would have been conducted to determine the effect of individual studies on the summary estimates.

Data analysis was performed using RevMan 5 (Version 5.4.1) (45) and the 'lme4' package in R (version 4.0.3) (46).

Results

Study Selection

Our literature search was updated through January 11th, 2022 (Schieda 2022 (47) was obtained after the search date after researchers in the area, who were known to be conducting studies for "preprint data", were contacted). We identified 2137 titles and abstracts for initial screening; 1852 articles were excluded after title and abstract review and 273 were excluded after full-text review. Ten unique articles were included in the review: 5 studies for the primary target condition comparing malignant and benign renal masses, 0 studies for the secondary target condition comparing aggressive and indolent masses, and 6 studies for the secondary outcome comparing ccRCC and other histological subtypes. The study by Baldari et al. (48) was included in both analyses. Study selection and the reasons for exclusion are detailed in Figure 2.1.

Study Characteristics

Table 2.1 provides the characteristics of the included studies. All studies were conducted between 2015 and 2022. In all studies, the reference standard and index test were performed before data collection. The review included a total of 1239 renal masses from approximately 1200 adult patients (one study did not report the number of patients). The number of masses in each study ranged from 18-454, with seven of the ten studies containing less than 100 masses. All studies used histopathological assessment as the reference standard to diagnose malignancy

or histological subtype. The specific histologic diagnosis of the masses in each study can be found in appendix 2.5.

Eight of the ten included studies evaluated multi-parametric MRI, with 6 of those studies using T2 weighted, chemical shift [T1 in phase and out of phase] and DCE sequences. Table 2.2 and appendix 6 and 7 provide a detailed description of the MRI sequences used in each study. The following studies used ccLS to diagnoses clear cell RCC from other renal masses: Steinberg 2021(18, 43), Canvasser 2017 (43) and Schieda 2022 (47).

Quality Assessment

The study-level results of the evaluation of risk of bias and applicability concerns, using the quality assessment of diagnostic accuracy studies-2 tool (QUADAS-2), are shown in Figure 2.2. The risk of bias in patient selection was unclear in seven studies because information about patient enrolment was not provided. The risk of bias in the index test was unclear in two studies and high in five studies because the diagnostic criteria were not adequately specified in the methods section (unclear risk of bias), or not reported (high risk of bias).

Applicability concerns for patient selection were high in four studies (proportion of cystic vs solid was not reported) and unclear in four studies (definition of "solid" was not specified). Applicability concerns regarding the index test were high in three studies and unclear in five studies since the timing of post-contrast phases were not specified in the former or did not resemble the recommendations of the "Society of Abdominal Radiology Disease Focused Panel on Renal Cell Carcinoma" in the latter (25).

Meta-Analysis Of Diagnostic Accuracy

The summary estimate for sensitivity in detecting malignant solid renal masses using bpMRI and mpMRI was 0.95 (95% CI, 0.77-0.99) and for specificity 0.63 (95% CI, 0.46-0.77) (Appendix 2.8). The forest plot of sensitivity and specificity for each included study is presented in appendix 2.8. The scatter of the study points is displayed on the SROC plot (Figure 2.3). The variance (tau-squared) of the random effects parameters is 2.35 and 0.28 for the sensitivity and specificity respectively. Overall, the sROC plot shows a high degree of variability in sensitivity and specificity.

No study had reported the diagnostic accuracy of bpMRI or mpMRI for aggressive versus indolent renal masses.

The summary estimate for sensitivity in differentiating clear cell RCC from other renal subtypes using bpMRI or mpMRI was 0.85 (95% CI, 0.77-0.90) and for specificity 0.77 (95% CI, 0.73-0.81) (Appendix 2.9). The forest plot of the sensitivity and specificity of each study is presented in appendix 2.9. The scatter of the study points is displayed on the SROC plot (Figure 2.4).

The variance (tau-squared) of the random effects parameters is 0.20 and 9.06×10^{-6} for the sensitivity and specificity, respectively. Overall, the sROC plot shows a moderate degree of variability in sensitivity and small degree of variability in the specificity.

The data from the studies that used ccLS scores to diagnose clear cell RCC are presented in the appendix (Figure 2.5, Appendix 2.10). A ccLS score of 5 had the highest specificity (Range: 0.92-0.93) and a ccLS score of 1 had the highest sensitivity (Range: 0.99-1.00). A meta-analysis for the ccLS scores was not conducted due to the limited number of studies.

The number of included studies in this review was insufficient to conduct the planned analyses for sources of variability for any of the objectives.

Discussion

This study aimed to synthesize the available evidence regarding the accuracy of bpMRI and mpMRI for diagnosis of solid renal masses. We observed a summary sensitivity of 95% at a specificity of 63% for differentiation of malignant from benign solid renal masses. For one of our secondary objectives, the ability of bpMRI or mpMRI to differentiate aggressive from indolent renal masses, we found no eligible studies. For our other secondary objective, diagnosis of clear cell RCC subtype, we observed a summary sensitivity of 85% for a specificity of 77%. Both meta-analyses demonstrate wide variability in study level point estimates, suggesting that the summary estimates may be influenced by other variables or be caused by random chance. The confidence region and prediction region are both quite large for the primary outcome, indicating that the true value of the mean accuracy, and the result of subsequent primary studies, could be

substantially different from the summary estimates determined in this study. Given the small number of included studies, identification of sources of variability was not possible.

In addition to the scarcity of published evidence on the outcomes of interest, there was high or unclear risk of bias in most included studies synthesized in this review, which limits our confidence in application of these results in routine clinical practice. The nature of the included primary studies is a potential source of variability, due to the heterogeneity with respect to the study populations (including definitions and proportions of solid versus cystic masses), imaging techniques (all studies used T2, 9/10 studies used DCE, and 6/10 studies used DWI and T1), and the diagnostic criteria used.

The available evidence on the diagnostic accuracy of MRI for differentiating benign from malignant solid renal masses is inconsistent. The results of previous studies are probably highly influenced by the study group and index test used, which are often not representative of the diagnostic accuracy of MRI in the clinical setting since they typically use nonconsecutive samples with preselected renal mass subtypes. For example, a study by Wilson et al. concluded that MRI was sensitive (83%) and highly specific (90%) in distinguishing lipid-poor AML, but the study sample was highly heterogenous (28). In this review, studies that included oncocytomas and other less common benign masses reported lower sensitivity with wider confidence intervals, compared with studies that contained only lipid-poor AML (28). In addition, studies that only included lipid-poor AML as benign masses were compared to a select group of renal mass subtypes, which is not reflective of clinical practice (28). From the studies included in our review, only De Silva 2021 showed good sensitivity and specificity, however selection of masses used in the study was unclear. Although the diagnostic accuracy of bpMRI and mpMRI is inconclusive, this review reinforces the need for consistent study design when evaluating the diagnostic accuracy of solid renal masses. The study sample requires a diverse, ideally consecutive, set of the various malignant and benign masses encountered in clinical practice, not solely focusing on a pre-selected subtype of masses.

Diagnostic tests to improve histology prediction may have clinical utility. Approximately 20% of small renal masses (<4cm) are benign and up to 80% are indolent (17, 49). For patients with small renal masses, age, sex, tumour size/location, symptoms, and smoking history are

associated with risk of malignancy and aggressive malignancy. However, the discriminative accuracy of these clinical and imaging characteristics is poor to moderate (17, 49-51). It is possible that the addition of mpMRI to existing clinical prediction models improves accuracy and allows more patients to avoid tumour biopsy and/or treatments.

Patients with metastatic RCC may also benefit from non-invasive determination of histology. Patients with metastases usually have large renal masses. The decision to perform a cytoreductive nephrectomy and the selection of systemic medical treatments are influenced by histologic subtypes where clear cell tumours are much more likely to respond to immunotherapy and tyrosine kinase inhibitors (52). An accurate and non-invasive assessment of clear cell histology could potentially avoid the need for biopsy and facilitate earlier initiation of systemic treatments.

In this review, three of the six studies evaluating the diagnostic accuracy of ccRCC had either a very small sample size (18 masses in Baldari 2015) (48) or minimal diversity in the sample population (less than two masses used in the comparator group in Nikpanah 2021 and Cupido 2017) (53, 54). These do not reflect the natural prevalence of clear cell RCC versus other tumors in clinical practice and influence the accuracy estimates. The other included studies showed moderate sensitivity and specificity, but these require validation of diagnostic criteria. Based on the few studies, at a high risk of bias, identified in this review, it is clear that more high-quality evidence is needed before mpMRI can be considered as part of a clinical prediction model.

Limitations

This systematic review has several limitations. Firstly, differences in the design of included studies increases the potential for selection bias. The imaging parameters and diagnostic criteria used have a direct effect on the quality of the index test, impacting the diagnostic accuracy of our index test. Moreover, the solid-to-cystic proportion, the number of different subtypes and the size of masses used could overestimate or underestimate the summary estimates for each study. Secondly, the estimates had a high degree of variability and a subgroup analysis could not be conducted to evaluate for sources of variability. Thirdly, several studies had a high or unclear risk of bias, often due to incomplete reporting of study details, creating uncertainty in

our synthesis of the literature. Lastly, spectrum effects could not be evaluated in this review due to the limited number of studies. Differences in the renal mass size or tumor grade may be associated with imaging characteristics, therefore the prevalence of tumour sizes and histology within each study may affect diagnostic performance.

Conclusion

In conclusion, the diagnostic accuracy of bpMRI and mpMRI for solid renal masses is unclear based on the available evidence. There are deficits in the availability and quality of evidence for the diagnosis of malignant pathology, aggressive pathology, and clear cell RCC among solid renal masses, Our results indicate substantial variability in the reported accuracy estimates for diagnosis of malignancy and ccRCC, due to unknown sources, however we observed differences in patient populations, definitions of solid masses, imaging techniques and diagnostic criteria between primary studies. No study has evaluated accuracy of MRI for diagnosis of aggressive versus indolent renal masses. Future primary studies can improve consistency and generalizability by emphasizing the use of clinically routine MRI techniques and evaluating a study population that is relevant to clinical practice. The implementation of reporting guidelines, standardized imaging techniques, and universally accepted definitions for solid, aggressive and indolent masses could improve study quality and the validity and applicability of final results.

Tables

Table 2.1. Characteristics of included studies

Study	Year of publication	Data Collection	Location	Number of included patients	Mean age (SD)[Min, Max]	Type of study sample	Number of masses	Index Test	Reference standard	Funding Source
Xu et al.a (55)	2021	After the index test and reference standard were applied	Single centre, United States	60	58 [18,87]	T2 isointense or hypointense on MRI	60	mpMRI	Biopsy or Resection	Not reported
Steinberg et al. (18)	2021	After the index test and reference standard were applied	Multi-centre, United States	434	60 (13) [18,91]	All solid renal masses	454	mpMRI	Surgery (387) + Biopsy (67)	NIH grants
Nikpanah et al. (53)	2021	After the index test and reference standard were applied	Single centre, United States	79	59.3	ccRCC + Oncocytoma	80 included in radiologist subset (of 243 total)	bpMRI	Histology, not specified	Intramural Research Programs of the Center for Cancer Research-National Cancer Institute and the National

										Institutes of Health Clinical Center
Goyal et al. (56)	2018	After the index test and reference standard were applied	Single centre, India	NR	42.7 [18,85]	All solid renal masses	90	mpMRI	Composite histology, typical imaging criteria, and follow-up imaging	None
Cupido et al. (54)	2017	After the index test and reference standard were applied	Single centre, Canada	50	NR	Only solid RCCs	50	bpMRI	Imaging follow-up (describe modalities and interval) + Surgical specimen (total vs. partial not reported)	None
Canvasser et al. (43)	2017	After the index test and reference standard were applied	Single centre, United States	110	57 (14)	cT1a renal masses	121	mpMRI	Partial or Total Nephrectomy	SPORE grant
Baldari et al. (48)	2015	After the index test and reference	Unclear/Not reported	17	68 [45,85]	All solid renal masses	18	mpMRI	Histology, sampling method not specified	Not reported

		standard were applied								
Xu et al.b (57)	2021	After the index test and reference standard were applied	Single centre, China	44	NR	All solid renal masses	44	bpMRI	Histopathology, not otherwise specified	Jiangsu Province Six First Project for High-level Health Professionals. Grant Number: LGY2019032 National Natural Science Foundation of China. Grant Number: 81401384
De Silva et al. (58)	2021	After the index test and reference standard were applied	Unclear/Not reported, Australia	72	NR	All solid renal masses	72	mpMRI	Histopathology, not otherwise specified (5 cases had pathognomic findings for AML; biopsy was considered unethical)	Not reported

Schieda et al. (47)	2022	After the index test and reference standard were applied	Multi-centre, Canada and the United States	241	60 (13)	Solid renal masses (≤4cm)	250	mpMRI	Surgical excision or image-guided percutaneous biopsy	Not reported

NR-Not reported

Table 2.2. MRI parameters used in included studies

Study	Index test sequence(s) applied	Magnetic field strength, T	Other phases used	Post-contrast phases		Enhancement Agent
				Corticomedullary phase timing (sec.)	Nephrographic phase timing (sec.)	
Xu 2021a	T2, T1 (in-phase, out-of-phase), DCE, DWI	1.5T/3.0T	No	Based on time running	40s after cortico-medullary phase	Gadopentetate dimeglumine or gadobutrol
Steinberg 2021	T2, T1 (in-phase, out-of-phase), DCE, DWI	1.5T/3.0T	No	NR	NR	NR
Nikpanah 2021	T2, T1, DCE	1.5T	No	20	70	Gadobutrol
Goyal 2018	T2, T1 (in-phase, out-of-phase), DCE, DWI	1.5T	No	NR	NR	Gadobenate dimeglumine
Cupido 2017	T2, DCE	1.5T	No	30	90	Gadobutrol
Canvasser 2017	T2, T1 (in-phase, out-of-phase), DCE	1.5T/3.0T	No	NR	NR	NR
Baldari 2015	T2, T1 (in-phase, out-of-phase), DCE	NR	No	40-55	90-120	Gadopentic acid
Xu 2021b	T2, DWI	3.0T	No	NA	NA	NA
De Silva 2021	T2, DCE, DWI	3.0T	No	NA	NA	NR
Schieda 2022	T2, T1 (in-phase, out-of-phase), DCE, DWI	1.5T/3.0T	No	30	100	Gadobenate dimeglumine

NA- Not applicable; NR-Not reported; DWI-Diffusion weighted imaging; DCE- Dynamic contrast enhancement

Figures

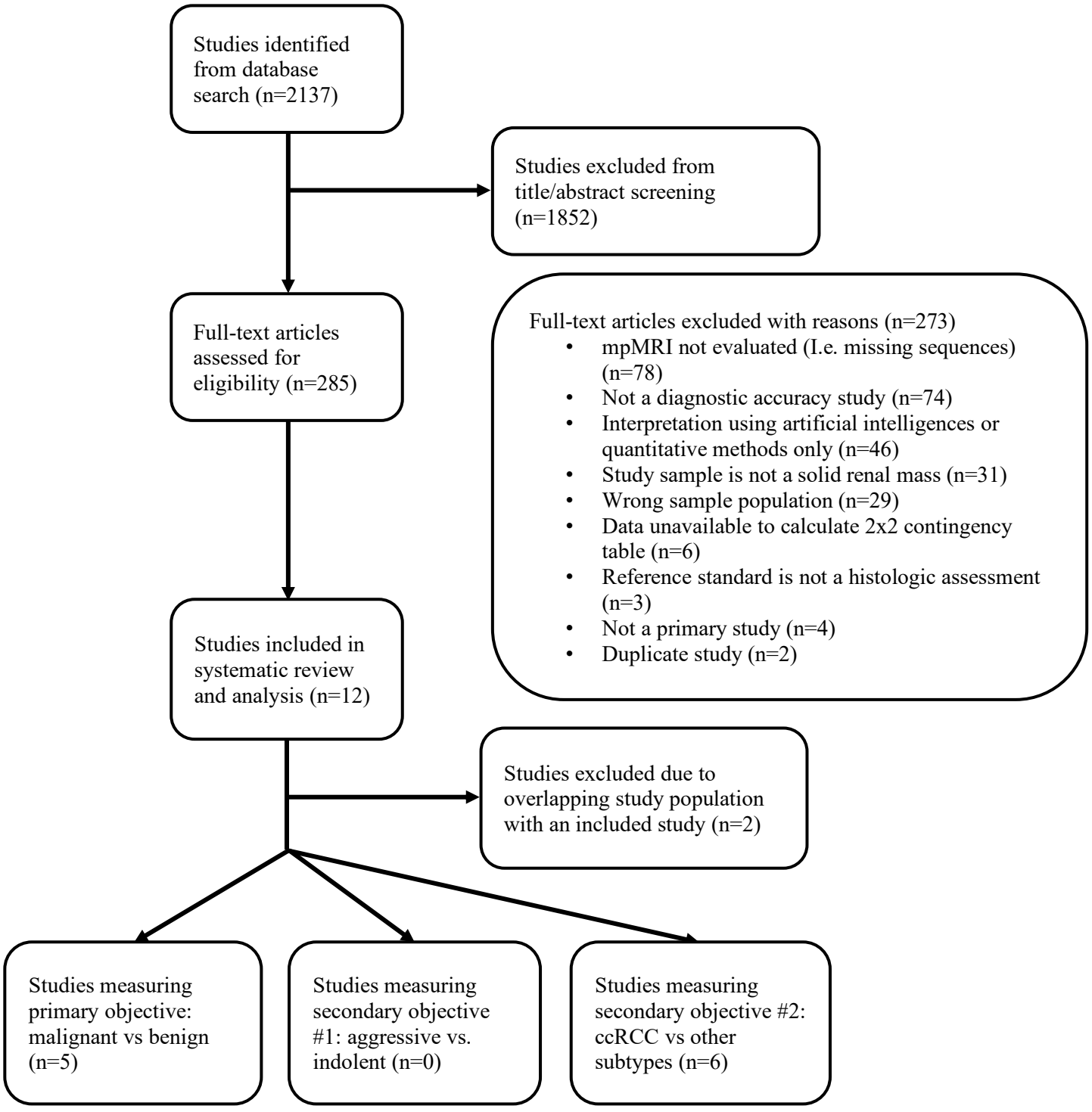


Figure 2.1. Flow diagram of study inclusion. The study by Baldari et al ([48]) was included in analyses for both the primary objective and secondary objective #2.

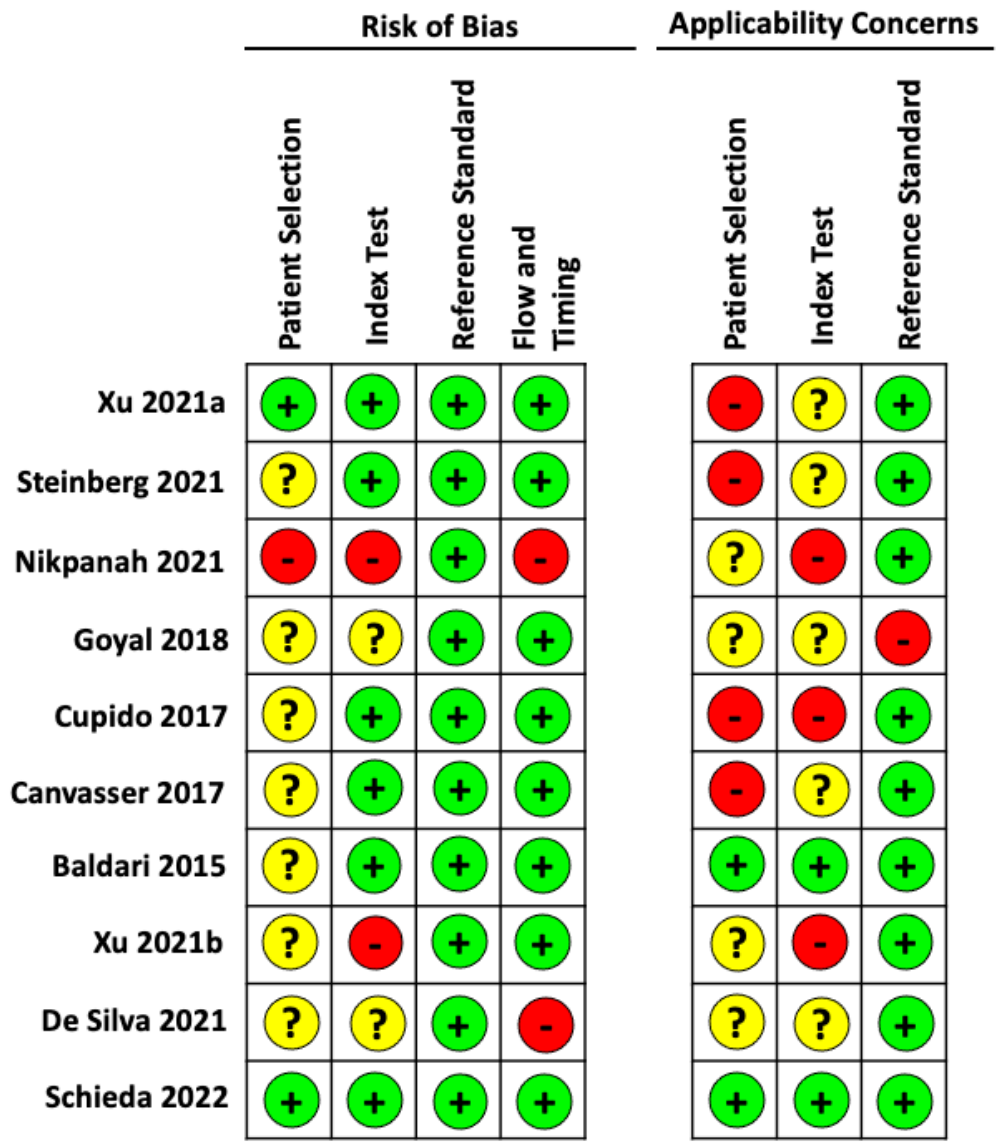


Figure 2.2. Risk of bias and applicability concerns for each study and domain using QUADAS-2. Risk of bias and applicability concerns were ranked as low (+) , unclear (?) or high (-).

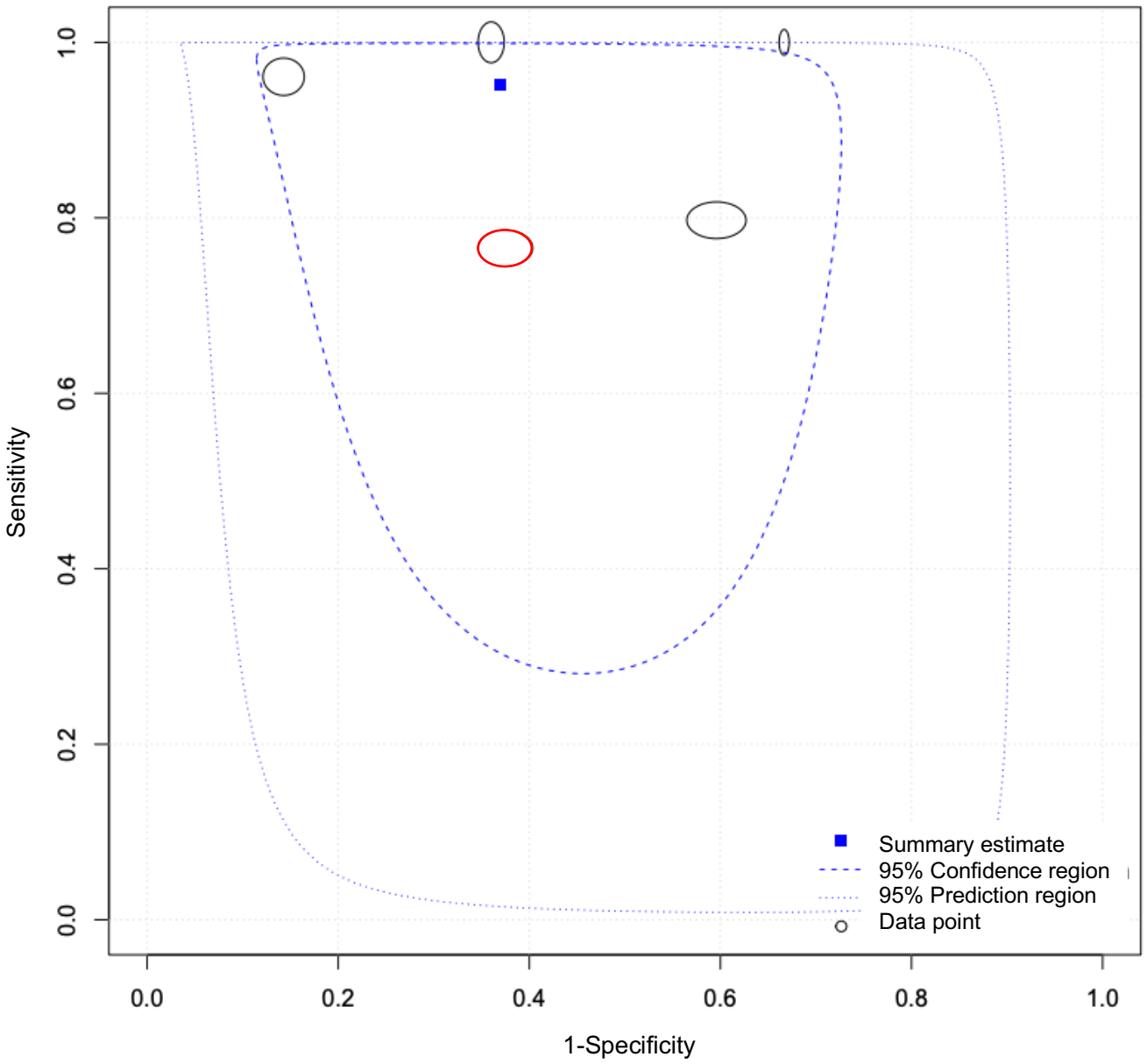


Figure 2.3. sROC plot for the primary target condition: malignant vs. benign solid renal masses (n=5). Summary point estimate, 95% confidence region and 95% predictive region were calculated using the bivariate random effects model based on the lme4 package in R. Individual study data is presented using the clear black circles, with the size of each data point representing the relative study weight. The solid blue square represents the summary point for all studies. The data points were color coded by index test (red: bpMRI; black: mpMRI). The figure was generated using the lme4 package in MetaDTA [35, 36].

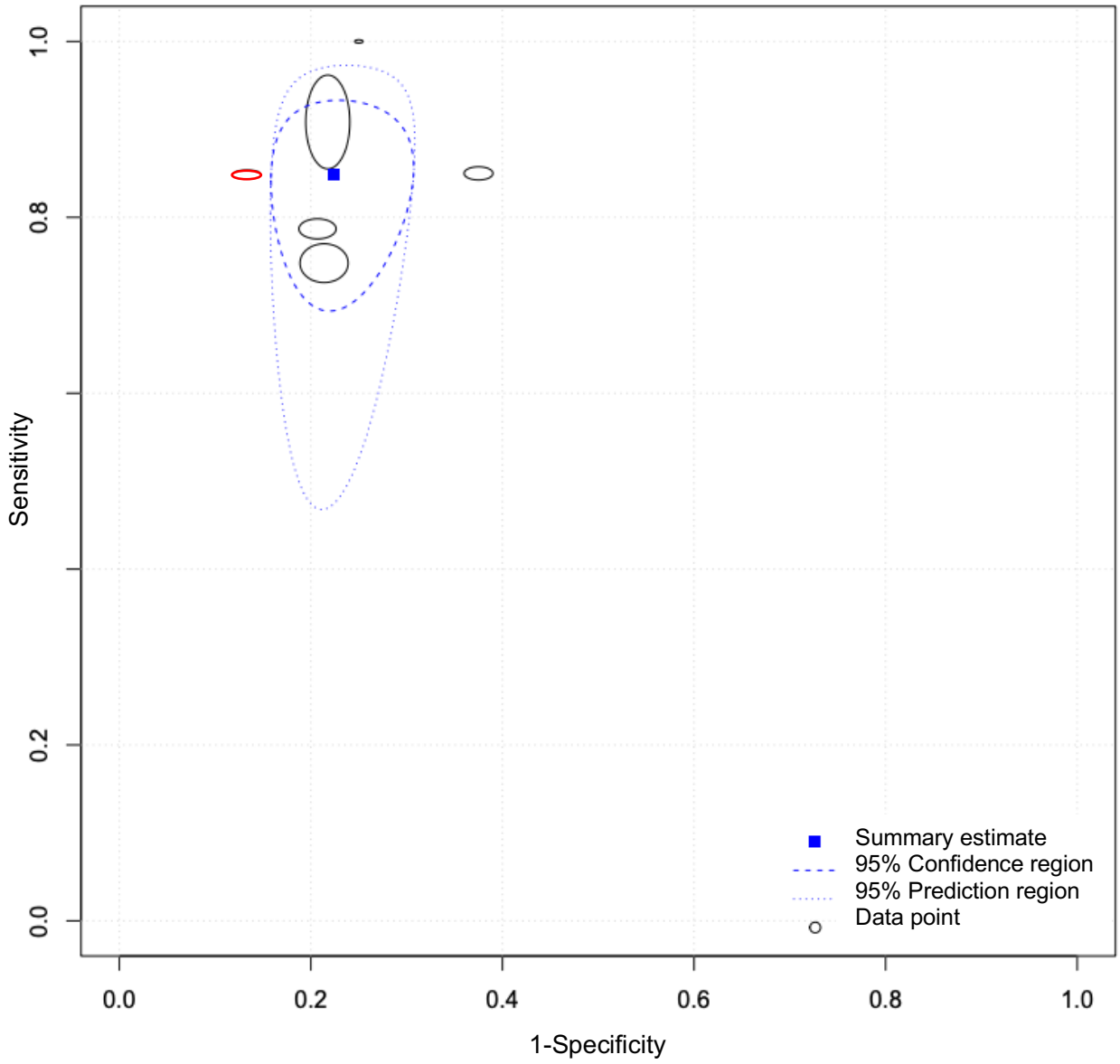


Figure 2.4. sROC plot studies for the secondary target condition: ccRCC vs other subtypes (n=6). Summary point estimate, 95% confidence region and 95% predictive region were calculated using the bivariate random effects model based on the lme4 package in R. Individual study data is presented using the clear black circles, with the size of each data point representing the relative study weight. The solid blue square represents the summary point for all studies. The data points were color coded by index test (red: bpMRI; black: mpMRI). The figure was generated using the lme4 package in MetaDTA [35, 36].For studies that used ccLS for diagnosis, masses were defined as ccRCC "positive" if they had score of 4 or 5 [18, 43], while other masses (ccLS 1-3) were classified as "negative."

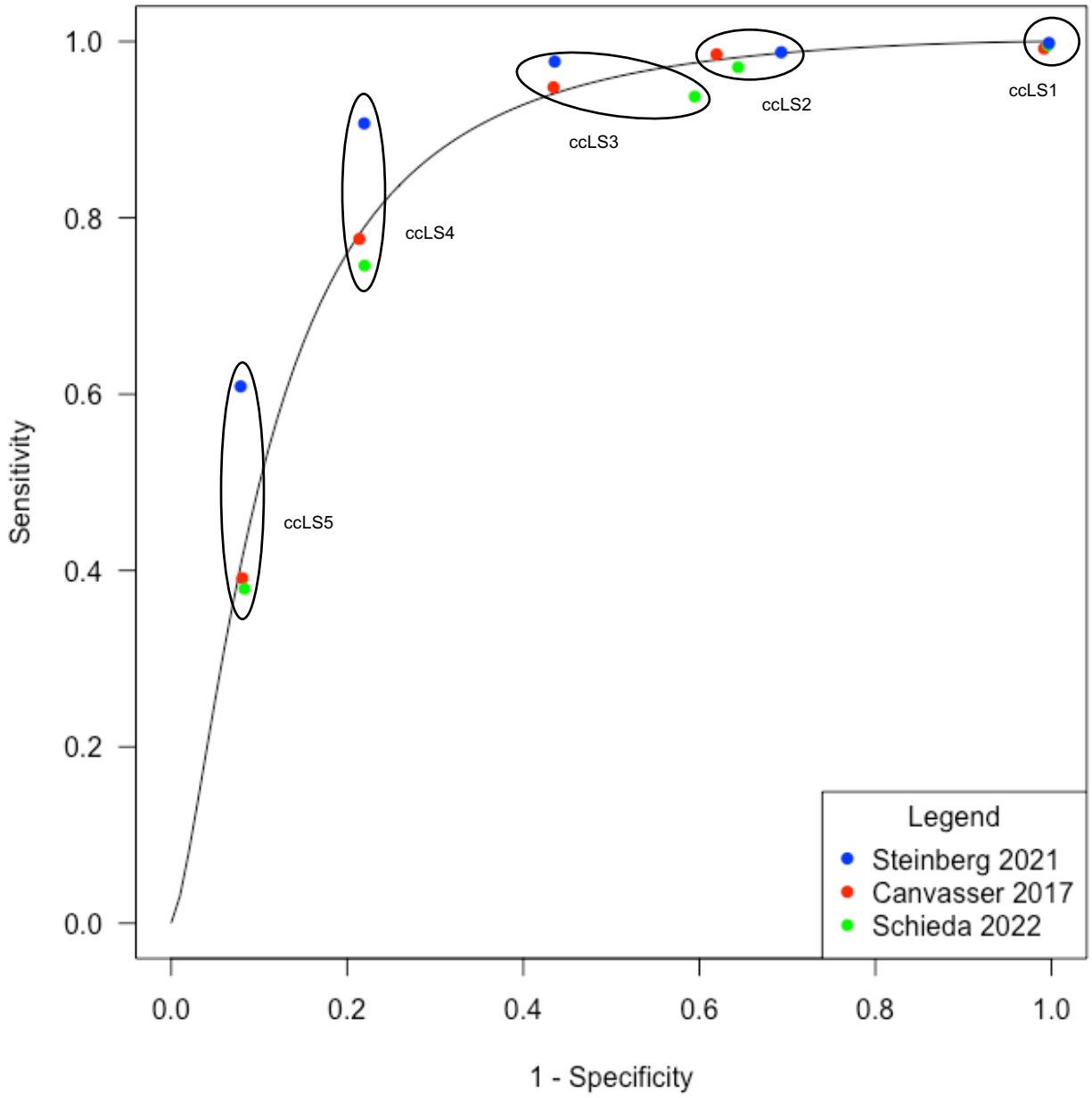


Figure 2.5. sROC curve of the sensitivity and specificity for each clear cell likelihood score (ccLS). Steinberg 2021 [18], Canavsser 2017 [43] and Schieda 2022 [47] used ccLS scores to diagnose ccRCC. Point estimates were calculated using the linear mixed effects model for modeling multiple threshold data, as described in Schneider et al. [40-42] The figure was generated using R.

References

1. Campbell S, Uzzo RG, Allaf ME, et al. Renal Mass and Localized Renal Cancer: AUA Guideline. J Urol. 2017;198:520-9.
2. Jayson M, Sanders H. Increased incidence of serendipitously discovered renal cell carcinoma. Urology. 1998;51:203-5.
3. Padala SA, Barsouk A, Thandra KC, et al. Epidemiology of Renal Cell Carcinoma. World J Oncol. 2020;11:79-87.
4. Hock LM, Lynch J, Balaji KC. Increasing incidence of all stages of kidney cancer in the last 2 decades in the United States: an analysis of surveillance, epidemiology and end results program data. J Urol. 2002;167:57-60.
5. Lightfoot N, Conlon M, Kreiger N, et al. Impact of noninvasive imaging on increased incidental detection of renal cell carcinoma. European urology. 2000;37:521-7.
6. May AM, Guduru A, Fernelius J, et al. Current Trends in Partial Nephrectomy After Guideline Release: Health Disparity for Small Renal Mass. Kidney Cancer. 2019;3:183-8.
7. Wong MCS, Goggins WB, Yip BHK, et al. Incidence and mortality of kidney cancer: temporal patterns and global trends in 39 countries. Scientific reports. 2017;7:15698.
8. Pierorazio PM, Johnson MH, Ball MW, et al. Five-year Analysis of a Multi-institutional Prospective Clinical Trial of Delayed Intervention and Surveillance for Small Renal Masses: The DISSRM Registry. European urology. 2015;68:408-15.
9. Jewett MAS, Mattar K, Basiuk J, et al. Active Surveillance of Small Renal Masses: Progression Patterns of Early Stage Kidney Cancer. European urology. 2011;60:39-44.
10. Chang SL, Cipriano LE, Harshman LC, Garber AM, Chung BI. Cost-effectiveness analysis of nephron sparing options for the management of small renal masses. J Urol. 2011;185:1591-7.
11. McAlpine K, Breau RH, Stacey D, et al. Development and acceptability testing of a patient decision aid for individuals with localized renal masses considering surgical removal with partial or radical nephrectomy. Urol Oncol. 2019;37:811.e1-.e7.
12. Chandrasekar T, Boorjian SA, Capitanio U, Gershman B, Mir MC, Kutikov A. Collaborative Review: Factors Influencing Treatment Decisions for Patients with a Localized Solid Renal Mass. European urology. 2021.

13. Remzi M, Ozsoy M, Klingler HC, et al. Are small renal tumors harmless? Analysis of histopathological features according to tumors 4 cm or less in diameter. J Urol. 2006;176:896-9.
14. Pahernik S, Ziegler S, Roos F, Melchior SW, Thuroff JW. Small renal tumors: correlation of clinical and pathological features with tumor size. J Urol. 2007;178:414-7; discussion 6-7.
15. Schlomer B, Figenshau RS, Yan Y, Venkatesh R, Bhayani SB. Pathological features of renal neoplasms classified by size and symptomatology. J Urol. 2006;176:1317-20; discussion 20.
16. Lee H, Lee JK, Kim K, et al. Risk of metastasis for T1a renal cell carcinoma. World Journal of Urology. 2016;34:553-9.
17. Bhindi B, Thompson RH, Lohse CM, et al. The Probability of Aggressive Versus Indolent Histology Based on Renal Tumor Size: Implications for Surveillance and Treatment. Eur Urol. 2018;74:489-97.
18. Steinberg RL, Rasmussen RG, Johnson BA, et al. Prospective performance of clear cell likelihood scores (ccLS) in renal masses evaluated with multiparametric magnetic resonance imaging. European radiology. 2021;31:314-24.
19. Silverman SG, Gan YU, Mortele KJ, Tuncali K, Cibas ES. Renal masses in the adult patient: the role of percutaneous biopsy. Radiology. 2006;240:6-22.
20. Marconi L, Dabestani S, Lam TB, et al. Systematic Review and Meta-analysis of Diagnostic Accuracy of Percutaneous Renal Tumour Biopsy. Eur Urol. 2016;69:660-73.
21. Patel D, Ghali F, Patel S, et al. Utilization of renal mass biopsy in patients with localized renal cell carcinoma: A population-based study utilizing the National Cancer Database. Journal of Clinical Oncology. 2020;38:662-.
22. Richard PO, Martin L, Lavallée LT, et al. Identifying the use and barriers to the adoption of renal tumour biopsy in the management of small renal masses. Can Urol Assoc J. 2018;12:260-6.
23. Menogue SR, O'Brien BA, Brown AL, Cohen RJ. Percutaneous core biopsy of small renal mass lesions: a diagnostic tool to better stratify patients for surgical intervention. BJU international. 2013;111:E146-51.
24. Patel HD, Johnson MH, Pierorazio PM, et al. Diagnostic Accuracy and Risks of Biopsy in the Diagnosis of a Renal Mass Suspicious for Localized Renal Cell Carcinoma: Systematic Review of the Literature. J Urol. 2016;195:1340-7.

25. Zhen Jane Wang, Matthew S. Davenport, Stuart G. Silverman, et al. MRI renal mass protocol v1.0: Society of Abdominal Radiology Disease Focused Panel on Renal Cell Carcinoma [

26. Ramamurthy NK, Moosavi B, McInnes MD, Flood TA, Schieda N. Multiparametric MRI of solid renal masses: pearls and pitfalls. Clin Radiol. 2014.

27. Kang SK, Zhang A, Pandharipande PV, Chandarana H, Braithwaite RS, Littenberg B. DWI for Renal Mass Characterization: Systematic Review and Meta-Analysis of Diagnostic Test Performance. AJR Am J Roentgenol. 2015;205:317-24.

28. Wilson MP, Patel D, Murad MH, McInnes MDF, Katlariwala P, Low G. Diagnostic Performance of MRI in the Detection of Renal Lipid-Poor Angiomyolipomas: A Systematic Review and Meta-Analysis. Radiology. 2020;296:511-20.

29. Group CDM. Handbook for DTA Reviews 2021 [Available from: https://methods.cochrane.org/sdt/handbook-dta-reviews.

30. Salameh JP, Bossuyt PM, McGrath TA, et al. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. BMJ. 2020;370:m2632.

31. Cohen JF, Deeks JJ, Hooft L, et al. Preferred reporting items for journal and conference abstracts of systematic reviews and meta-analyses of diagnostic test accuracy studies (PRISMA-DTA for Abstracts): checklist, explanation, and elaboration. BMJ. 2021;372:n265.

32. Moch H, Cubilla AL, Humphrey PA, Reuter VE, Ulbright TM. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part A: Renal, Penile, and Testicular Tumours. Eur Urol. 2016;70:93-105.

33. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529-36.

34. Network E. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement 2018 [Available from: https://www.equator-network.org/reporting-guidelines/prisma-dta/.

35. Patel A, Cooper N, Freeman S, Sutton A. Graphical enhancements to summary receiver operating characteristic plots to facilitate the analysis and reporting of meta-analysis of diagnostic test accuracy data. Res Synth Methods. 2021;12:34-44.

36. Freeman SC, Kerby CR, Patel A, Cooper NJ, Quinn T, Sutton AJ. Development of an interactive web-based tool to conduct and interrogate meta-analysis of diagnostic test accuracy studies: MetaDTA. BMC Med Res Methodol. 2019;19:81.
37. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005;58:982-90.
38. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. J Clin Epidemiol. 2006;59:1331-2; author reply 2-3.
39. McGrath TA, McInnes MDF, Langer FW, Hong J, Korevaar DA, Bossuyt PMM. Treatment of multiple test readers in diagnostic accuracy systematic reviews-meta-analyses of imaging studies. Eur J Radiol. 2017;93:59-64.
40. Steinhauser S, Schumacher M, Rücker G. Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. BMC Med Res Methodol. 2016;16:97.
41. Schneider A, Linde K, Reitsma JB, Steinhauser S, Rücker G. A novel statistical model for analyzing data of a systematic review generates optimal cutoff values for fractional exhaled nitric oxide for asthma diagnosis. J Clin Epidemiol. 2017;92:69-78.
42. Vali Y, Lee J, Boursier J, et al. Enhanced liver fibrosis test for the non-invasive diagnosis of fibrosis in patients with NAFLD: A systematic review and meta-analysis. J Hepatol. 2020;73:252-62.
43. Canvasser NE, Kay FU, Xi Y, et al. Diagnostic Accuracy of Multiparametric Magnetic Resonance Imaging to Identify Clear Cell Renal Cell Carcinoma in cT1a Renal Masses. The Journal of urology. 2017;198:780-6.
44. Pedrosa I, Cadeddu JA. How We Do It: Managing the Indeterminate Renal Mass with the MRI Clear Cell Likelihood Score. Radiology. 2022;302:256-69.
45. *Review Manager (RevMan) [Computer program]. Version 5.4.1, The Cochrane Collaboration, 2020.*
46. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. Vienna, Austria: R Foundation for Statistical Computing; 2020 [

47. Schieda N, Davenport MS, Silverman SG, et al. Multicenter Evaluation of Multiparametric MRI Clear Cell Likelihood Scores in Solid Indeterminate Small Renal Masses. Radiology. 2022;303:590-9.
48. Baldari D, Capece S, Mainenti PP, et al. Comparison between computed tomography multislice and high-field magnetic resonance in the diagnostic evaluation of patients with renal masses. Quantitative imaging in medicine and surgery. 2015;5:691-9.
49. Golan S, Eggener S, Subotic S, et al. Prediction of renal mass aggressiveness using clinical and radiographic features: a global, multicentre prospective study. BJU Int. 2016;117:914-22.
50. Frank I, Blute ML, Cheville JC, Lohse CM, Weaver AL, Zincke H. Solid renal tumors: an analysis of pathological features related to tumor size. J Urol. 2003;170:2217-20.
51. Lane BR, Babineau D, Kattan MW, et al. A preoperative prognostic nomogram for solid enhancing renal tumors 7 cm or less amenable to partial nephrectomy. J Urol. 2007;178:429-34.
52. Canil C, Kapoor A, Basappa NS, et al. Management of advanced kidney cancer: Kidney Cancer Research Network of Canada (KCRNC) consensus update 2021. Can Urol Assoc J. 2021;15:84-97.
53. Nikpanah M, Xu Z, Jin D, et al. A deep-learning based artificial intelligence (AI) approach for differentiation of clear cell renal cell carcinoma from oncocytoma on multi-phasic MRI. Clin Imaging. 2021;77:291-8.
54. Cupido BD, Sam M, Winters SD, et al. A practical imaging classification for the non-invasive differentiation of renal cell carcinoma into its main subtypes. Abdominal radiology (New York). 2017;42:908-17.
55. Xu HS, Balcacer P, Zhang Z, et al. Characterizing T2 iso- and hypo-intense renal masses on MRI: Can templated algorithms improve accuracy? Clinical imaging. 2021;72:47-54.
56. Goyal A, Sharma R, Bhalla AS, Gamanagatti S, Seth A. Comparison of MDCT, MRI and MRI with diffusion-weighted imaging in evaluation of focal renal lesions: The defender, challenger, and winner! The Indian journal of radiology & imaging. 2018;28:27-36.
57. Xu Q, Zhu Q, Liu H, et al. Differentiating Benign from Malignant Renal Tumors Using T2- and Diffusion-Weighted Images: A Comparison of Deep Learning and Radiomics Models Versus Assessment from Radiologists. J Magn Reson Imaging. 2021.

58. De Silva S, Lockhart K, Aslan P, et al. The differentiation of benign from malignant solid renal masses with multi-parametric MRI: A retrospective study and proposed classification scheme. European Urology Open Science2021. p. S310.

Chapter 3. Association of LI-RADS CT/MRI Ancillary Features with Hepatocellular Carcinoma (HCC) and Malignancy: Systematic Review and Individual Participant Data Meta-Analysis

List of authors

1. Haben Dawit, BSC. Clinical Epidemiology Program, Ottawa Hospital Research Institute. School of Epidemiology, Public Health and Preventative Medicine, Faculty of Medicine, University of Ottawa.
2. Eric Lam, MSc. The Ottawa Hospital Research Institute, Ottawa, ON.
3. Matthew DF McInnes MD Ph.D. FRCPC. University of Ottawa Department of Radiology. Clinical Epidemiology Program, Ottawa Hospital Research Institute.
4. Christian B. van der Pol MD FRCPC, Juravinski Hospital, and Cancer Centre, Hamilton Health Sciences, McMaster University.
5. Mustafa R. Bashir, MD. Department of Radiology, Duke University Medical Center, Durham, NC, USA; Center for Advanced Magnetic Resonance Development, Duke University Medical Center, Durham, NC, USA.
6. Jean-Paul Salameh, MSc. Faculty of Health Sciences, Queen's University. The Ottawa Hospital Research Institute Clinical Epidemiology Program
7. Brooke Levis, Ph.D. Centre for Clinical Epidemiology, Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, H3T 1E2.
8. Claude B. Sirlin, MD, Liver Imaging Group, Department of Radiology, UC San Diego.
9. Victoria Chernyak, MD MS. Department of Radiology, Memorial Sloan Kettering Cancer Center, NYC, NY.
10. Sang Hyun Choi, MD PhD. Department of Radiology and Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea.
11. So Yeon Kim, MD. Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Songpa-Gu, Seoul, Republic of Korea.
12. Tyler J. Fraum, MD. Mallinckrodt Institute of Radiology, Washington University School of Medicine, St Louis, MO.

13. An Tang, MD, MSc. Department of Radiology, Centre Hospitalier de l'Université de Montréal, Montréal, Québec, Canada.
14. Hanyu Jiang, MD. Department of Radiology, West China Hospital, Sichuan University, Chengdu, China.
15. Bin Song, MD. Department of Radiology, Sichuan University West China Hospital, Chengdu, Sichuan, China.
16. Jin Wang, MD. Department of Radiology, The Third Affiliated Hospital, Sun Yat-sen University (SYSU), Guangzhou, Guangdong, P.R. China.
17. Stephanie R. Wilson, MD. University of Calgary, Foothills Medical Centre, Calgary, Canada.
18. Heejin Kwon, MD. Department of Radiology, Dong-A University Hospital, Dong-A University College of Medicine, Seogu, Busan, Republic of Korea.
19. Andrea S. Kierans, MD. Weill Cornell Medical Center, New York City, NY, USA.
20. Ijin Joo, MD. Department of Radiology, Seoul National University College of Medicine, Jongno-gu, Seoul, Korea.
21. Maxime Ronot, MD PhD. Department of Radiology, APHP.Nord, Hopital Beaujon, Clichy & Université Paris Cité, CRI UMR 1149, Paris, France.
22. Ji Soo Song, MD. Department of Radiology, Chonbuk National University Medical School and Hospital, Deokjin-gu, Jeonju, Chonbuk, South Korea.
23. Joanna Podgórska, MD. 2nd Radiology Department, Warsaw Medical University, Warsaw, Poland.
24. Grzegorz Rosiak, MD PhD. 2nd Radiology Department, Warsaw Medical University, Warsaw, Poland.
25. Zhen Kang, MD. Department of Radiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Hubei, Wuhan, PR China.
26. Brian C. Allen, MD. Department of Radiology, Duke University Medical Center, Durham, NC.
27. Andreu F Costa, MD MSc FRCPC *(Corresponding Author).* Department of Diagnostic Radiology, Queen Elizabeth II Health Sciences Centre and Dalhousie University, Victoria General Building, 3rd floor, 1276 South Park Street, Halifax, Nova Scotia B3H 2Y9.

Preface

Manuscript Status

Chapter 3 was for publication to Radiology on June 15th, 2023.

Objective

The objective of this IPD systematic review and meta-analysis was the evaluate the association of the Liver Imaging Reporting and Data System (LI-RADS) AFs with hepatocellular carcinoma (HCC) and malignancy.

Contributions of authors

- Conception of the study protocol: Costa AF, Dawit H, van der Pol CB, Bashir MR, McInnes MDF
- LI-RADS IPD contributing authors: Choi SH, Kim SY, Fraum TJ, Tang A, Jiang H, Song B, van der Pol CB, Wang J, Wilson SR, Kwon H, Kierans AS, Joo I, Ronot M, Song JS, Podgórska J, Rosiak G, Kang Z, Allen BC
- Quality assessment: van der Pol CB, Salameh JP, Costa AF
- Data cleaning and validation: Dawit H, Lam E, Costa AF,
- Data storage (REDCap): Dawit H, Lam E
- Data analysis and interpretation: Dawit H, Lam E
- Drafting of the manuscript: Dawit H, Costa AF
- Revision of the manuscript: Lam E, van der Pol CB, Bashir MR, Salameh JP, Levis B, Chernyak V, Sirlin CB, McInnes MDF
- Research guidance and support: Costa AF, van der Pol CB, Bashir MR, McInnes MDF

Protocol registration

The protocol is registered on the Open Science Framework (https://osf.io/rpg8x)

Ethics approval

IRB approval was received from the Ottawa Hospital

Appendices

Supplementary material is available at the end of this thesis in appendix 3.1-3.12. Please refer to list of appendices on page VIII for specific page numbers.

Funding

The project was supported by the Joan Sealy Trust for Cancer Research, CIHR Project Grant and the Radiological Society of North America (RSNA) Research Scholar Grant. Dawit H was supported by the Queen Elizabeth II Graduate Scholarship in Science and Technology (QEII-GSST) Program and the Graduate Studentship for advanced training (OHRI)

Abstract

Background: The independent contribution of each Liver Imaging Reporting and Data System (LI-RADS) ancillary feature (AF) is not established.

Purpose: To evaluate the association of each AF with hepatocellular carcinoma (HCC) and malignancy, while adjusting for major features through an individual participant data (IPD) meta-analysis.

Materials and Methods: MEDLINE, Embase, Cochrane Central Register of Controlled Trials, and Scopus databases were searched from 2014–2022 for studies evaluating the diagnostic accuracy of CT and MRI for HCC, using LI-RADS v2014/2017/2018). Using a one-step approach, IPD across studies were pooled. Adjusted odds ratios (ORs) and 95% confidence intervals (CIs) were derived from multivariable logistic regression models of each AF combined with major features except threshold growth, which was excluded because of infrequent reporting. Liver observation clustering was addressed at the study and patient levels through random intercepts. Risk of bias was assessed using Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2). (Open Science Framework protocol: https://osf.io/rpg8x).

Results: Twenty studies comprising 3091 observations (2456 patients) were included. 89% (8/9) AFs favoring malignancy were associated with malignancy and/or HCC, 80% (4/5) AFs favoring HCC were associated with HCC, and 57% (4/7) AFs favoring benignity were negatively associated with HCC and/or malignancy. Nonenhancing capsule (OR: 3.50[95% CI=1.53-8.01]) and subthreshold growth (OR: 3.21[95% CI=1.26-8.17]) had the strongest association with HCC. Restricted diffusion (OR: 14.45[95% CI=9.82-21.27]) and mild-moderate T2 hyperintensity (OR: 10.18[95% CI=7.17-14.44]) had the strongest association with malignancy. The strongest negative association with HCC were parallels blood pool (OR: 0.07[95% CI=0.01-0.49]) and marked T2 hyperintensity (OR: 0.18[95% CI=0.07-0.45]). Seventeen studies (85%) had a high risk of bias.

Conclusion: Most LI-RADS AFs are independently associated with HCC, malignancy, or benignity as intended, when adjusting for major features. Our results highlight areas for refining

LI-RADS. Future LI-RADS revisions would be better informed with more research using standardized methodology.

Summary Statement: 89% (8/9) AFs favoring malignancy were associated with malignancy and/or HCC, 80% (4/5) AFs favoring HCC were associated with HCC, and 57% (4/7) AFs favoring benignity were associated with benignity.

Key Results:

- Apart from fat sparing in solid mass, all AFs favoring malignancy in general showed an independent association with HCC or malignancy.
- Apart from blood products in mass, all AFs favoring HCC in particular showed an independent association with HCC.
- Of AFs favoring benignity, size stability >2 years, size reduction, and undistorted vessels did not show an independent association with benignity

Abbreviations:

LI-RADS: Liver imaging reporting and data system; CT: Computed tomography; MRI: Magnetic resonance imaging; AF: Ancillary feature; HCC: hepatocellular carcinoma

Introduction

The American College of Radiology Liver Imaging Reporting and Data System (LI-RADS, LR) (1) is a standardized guideline for performing, interpreting and reporting imaging of patients at risk for hepatocellular carcinoma (HCC). LI-RADS is integrated with the American Association for the Study of Liver Diseases (AASLD) 2018 and 2023 HCC clinical practice guidance (2, 3). LI-RADS provides several advantages, including standard technical parameters for CT, MRI, and contrast-enhanced ultrasound (CEUS) (4) of the liver, a consistent terminology for describing imaging findings to avoid ambiguity with respect to HCC risk (5), and a framework for interpreting major and ancillary imaging findings to provide ordinal categories reflecting a probability of HCC and overall malignancy (6).

Although LI-RADS has undergone four major updates in response to growing evidence (7), knowledge gaps remain. One identified by the LI-RADS Evidence and Research Working Group is the contribution of ancillary features (AFs) towards LI-RADS diagnostic performance (8). Whereas major features have been extensively investigated, including in systematic reviews (9, 10), evaluations of inter-reader variability (11, 12), and a recent individual participant data (IPD) meta-analysis (13), there is little evidence on the diagnostic value of AFs independent of major features. AFs are imaging findings that favor either malignancy in general, HCC in particular, or benignity. AFs can be used to upgrade or downgrade the LI-RADS category assigned to an observation (1). When AFs are applied, the LI-RADS category may only be adjusted by one, irrespective of the number of AFs observed, and AFs cannot be used to up-categorize an observation from LR-4 (probable HCC) to LR-5 (definite HCC) (1, 14).

The use of AFs is controversial and adds complexity to the LI-RADS framework (8). Unlike major features, AFs are optional (i.e., applied at the radiologist's discretion). This option therefore represents a potential source of inter-reader variability because AFs can modify the LI-RADS category in 10-35% of observations (11, 14). Because several AFs are more commonly or exclusively evident on MRI (11), AFs also contribute to inter-modality variability. The diagnostic importance of each individual AF is not well established. Although AFs are weighted similarly with respect to LI-RADS category adjustments, studies have shown stronger associations between some AFs and malignancy, such as restricted diffusion

and hepatobiliary phase hypointensity (8). Therefore, the diagnostic value of AFs in the context of LI-RADS application requires further investigation.

The primary objective of this IPD meta-analysis was to evaluate the association of each AF with HCC and malignancy, after adjusting for major features. A secondary objective was to assess the proportion of observations that exhibited each AF.

Methods

This study analyzed participants included in the LI-RADS IPD, which was previously synthesized for an IPD meta-analysis on the association of LI-RADS major features with HCC (13). In this study, we assessed the association of LI-RADS AFs with HCC and malignancy using a partially overlapping, but expanded, dataset from Ref. (13). The analyses in the current work do not overlap with analyses previously reported from this data set. Studies derived from the database received IRB approval from The Ottawa Hospital. The study protocol is available on the Open Science Framework (https://osf.io/rpg8x). No personal health information was shared with project team members. This study adhered to the methodologic guidance from the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (15, 16). Study reporting followed Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data (PRISMA-IPD) and PRISMA for Diagnostic Test Accuracy (DTA) statements (17, 18).

Study Eligibility Criteria

Studies eligible for inclusion evaluated any CT, MRI, and/or CEUS LI-RADS AFs in association with HCC and overall malignancy. Other inclusion criteria were reported previously (13). Only participants at high risk of HCC were included, as defined by LI-RADS (cirrhosis, chronic hepatitis B viral infection, or current or prior HCC) (1). Participants in studies that did not meet the inclusion criteria were excluded. Studies conformed with LI-RADS technical guidelines for multiphasic contrast-enhanced CT, MRI, and US (4). Imaging features of observations were reported according to LI-RADS versions 2014, 2017, or 2018. Previously treated observations were excluded. A preferred composite reference standard was previously established to assess risk of bias (Appendix 3.1) (13).

Study Selection and Data Collection

The search strategy, study selection, and data collection process were reported previously (13) and are provided in appendix 3.2. An updated search was performed to include studies published up to January 2022, using the same search strategy as in (13). Data sharing agreements were obtained and de-identified data were transferred to an encrypted Research Electronic Data Capture (REDCap) database at The Ottawa Hospital (19, 20). Data from each study were validated by a research assistant (EL), graduate student (HD), and board-certified abdominal radiologist with experience in applying LI-RADS (AFC). Corresponding study authors were contacted when clarifications regarding data structure or integrity were required. For example, most studies required clarification on how major and ancillary features that require a prior examination (e.g., threshold growth [TG], subthreshold growth, and size stability > 2 years) were encoded, and whether this was compliant with the IPD data dictionary (Appendix 3.3). LI-RADS major features were classified as present or absent; in the case of TG, not applicable was also an option. Ancillary features were classified as present, absent, or not evaluated. For some AFs, not applicable was also an option (eg. AFs requiring a prior imaging examination, such as subthreshold growth, or AFs requiring a hepatobiliary agent). Imaging features that were not applicable were excluded from the analysis.

Risk of Bias and Applicability

Previously, a Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool (21) customized for studies related to LI-RADS was applied to assess for risk of bias (13) (Appendix 3.4). QUADAS-2 evaluates sources of bias along four domains: patient selection, index test, reference standard, and flow and timing. Risk of bias and applicability assessments were performed in duplicate and independently by two authors (CBvdP, JPS). Assessments were based primarily on the published study. Following validation of IPD, the assessments were reviewed to ensure there were no significant changes to the judgments (AFC).

Statistical Analysis

Using a one-step IPD meta-analysis approach, the IPD across studies were pooled, and models were constructed as follows. Adjusted odds ratios (ORs) and 95% confidence intervals (CIs) were derived for the association of each AF with HCC in particular (hereafter referred to as HCC) and malignancy in general (hereafter referred to as malignancy), in combination with the

LI-RADS major features. An initial study protocol planned for multivariable analyses to include TG, however, due to limited sample sizes of data with TG, multivariable analyses were conducted in two ways. In the primary analysis, TG was excluded as a major feature to maximize sample size and robustness of the models. In this analysis, TG and subthreshold growth were combined as an ancillary feature to determine the association between any growth and HCC or malignancy. As a secondary analysis, TG was included with the other four major features in accordance with LI-RADS. Liver observation clustering was addressed at the study and patient levels through random intercepts. Collinearity between variables was assessed by calculating the variance inflation factor, the tolerance statistic, and eigenvalues. I^2 statistic for heterogeneity, and the estimate of the between-study variance, tau^2, were obtained through a maximum-likelihood estimator. Due to a limited number of studies at low risk of bias, a sensitivity analysis was not conducted.-Publication bias was not assessed per contemporary guidance for DTA systematic reviews (18).

The level of significance was set at $p<0.05$ and calculated using a Wald test. All analyses were performed independently by HD and EL using the glmer function in the Lme4 package in R (R Core Team, version 4.0.0; R Foundation for Statistical Computing) (22).

Results

Study Selection and Characteristics

The study flow diagram is presented in Figure 3.1. Seventeen studies met inclusion criteria from the Jan 2014–Sept 2019 search performed in Ref. (13). Five studies were included from the updated search between Jan 2020–Jan 2022, for a total of 22 studies: 5 studies evaluating CT, 19 studies evaluating MRI and 2 studies evaluating CEUS. Four studies evaluated AFs using both CT and MRI (23-26). A summary of the characteristics of the 22 included studies is provided in Table 3.1. Due to limited sample size (185 observations), multivariable analysis of CEUS AFs was not conducted. The distribution of CEUS AFs is reported in appendix 3.5.

In total, 3091 unique CT/MRI liver observations from 2456 adult patients (mean age, 59±11 years, 75.3% male) were included. All studies relied on readers evaluating images, not the original reports, and used pathology or a composite reference standard to diagnose malignancy and HCC (Table 3.2).

Quality Assessment

Figure 3.2 shows the study-level results of the risk of bias and applicability concerns. 5/22 studies were considered at low risk of bias and 20/22 studies had low concerns regarding applicability. The flow and timing domain was the most common source of bias because many studies had unclear or inappropriate timing between the index test and the reference standard. In addition, risk of bias in the patient and observation selection domain was high in most studies due to study design (i.e., case-control design) or patient enrollment methods (i.e., non-consecutive, non-random, or select patients).

Association of ancillary features with HCC and malignancy

Primary analysis

Results of the primary multivariable analysis (TG excluded) are provided in Table 3.3, Figure 3.3, Figure 3.4, Appendix 3.6 and Appendix 3.7. The number of observations, patients, and studies included in each model are indicated in Table 3.3. The heterogeneity within studies (I^2) and between study-variance (tau^2) in each random effects model are presented in Table 3.4. The I^2 value for 18/22 models assessing HCC and 17/22 models assessing malignancy was greater than 90%.

Eight of nine AFs favoring malignancy in general were independently associated with HCC and/or malignancy. Corona enhancement was associated with malignancy only (OR:4.89[95% CI=1.71-14.01]). Ultrasound visibility as a discrete nodule (OR:3.18[95% CI=1.77-5.73]) and iron sparing in a solid mass (OR:2.63[95% CI=1.30-5.31]) were associated with HCC only. Fat sparing in a solid mass was not associated with HCC (OR:0.81[95% CI=0.44-1.49]) or malignancy (OR:0.79[95% CI=0.30-2.09]). The growth imaging feature, which combined TG and subthreshold growth, was associated with both HCC and malignancy.

Four of five AFs favoring HCC in particular were independently associated with HCC: mosaic architecture, nonenhancing capsule, nodule-in-nodule, and fat in mass. Mosaic architecture was associated with HCC (OR:2.53[95% CI=1.58-4.04]) and malignancy (OR:3.26[95% CI=1.25-8.51]). Nonenhancing capsule (OR:3.50[95% CI=1.53-8.01]), nodule-in-nodule (OR:2.15[95% CI=1.15-4.04]), and fat in mass (more than adjacent liver) (OR:1.44[95%

CI=1.03-2.01]) were associated with HCC only. Blood products in mass was not associated with HCC (OR:1.28[95% CI=0.83-1.96]) or malignancy (OR:2.31[95% CI=0.81-6.57].

Three of seven AFs favoring benignity were associated with decreased odds of HCC and malignancy (Appendix 3.6, Appendix 3.7): parallels blood pool, marked T2 hyperintensity, and hepatobiliary phase isointensity. Iron in mass (more than liver) (OR:0.07[95% CI=0.01-0.44]) was negatively associated with malignancy. Size stability (>2 years), size reduction, and undistorted vessels were not independently associated with either HCC or malignancy.

Secondary analysis

Results from the secondary analysis are included in appendix 3.8.

Proportions of observations that exhibited an AF

The proportions of observations that exhibited each AF are listed in Table 3.3 (primary analysis) and appendix 3.9 (secondary analysis). In the primary analysis for HCC, the average positivity rates were as follows: 47.3% (range, 3.9%-88.3%) for AFs favoring malignancy in general; 9.8% (range, 3.9%-15.4%) for AFs favoring HCC in particular; and 3.3% (range, 0.9%-8.7%) for AFs favoring benignity. Size reduction was the least common AF (3/348 observations, 0.9%), and ultrasound visibility was the most common AF (798/904, 88.3%).

Discussion

Our study demonstrated that most AFs are associated with malignancy in general, HCC in particular, or benignity as intended by LI-RADS after adjusting for major features other than threshold growth. Of the 8/9 AFs favoring malignancy in general that were associated with malignancy and/or HCC, the strongest associations were found with restricted diffusion (OR:14.5) and mild-moderate T2 hyperintensity (OR:10.2). Of the AFs favoring HCC specifically, 4/5 were associated with HCC (the exception being blood products in mass), and the strongest association with HCC was found with nonenhancing capsule (OR:3.5). Of the AFs favoring benignity, 4/7 were negatively associated with malignancy. Parallels blood pool (OR:0.02) and marked T2 hyperintensity (OR:0.06) had the strongest association with benignity.

A few AFs did not show independent associations with malignancy or benignity according to their intended use in LI-RADS. Of AFs favoring malignancy in general, US visibility as a discrete nodule was associated with HCC but not malignancy. US visibility had the highest positivity rate in our study (88.3% and 91.1% for HCC and malignancy, respectively), likely because US is used for HCC surveillance and because several studies used US visibility as a method to identify patient cohorts (25, 27). Fat sparing in a solid mass was not independently associated with either malignancy or HCC, and iron sparing in a solid mass was associated with HCC but not malignancy. Of note, these AFs had the lowest positivity rates of the AFs favoring malignancy in general (3.6% and 6.4% for malignancy, respectively). Of the AFs favoring HCC specifically, blood products in mass was not associated with HCC or malignancy. The positivity rate of this AF was low (9.2% for HCC), although other AFs favoring HCC specifically had similar if not lower positivity rates. Of AFs favoring benignity, three did not show a negative association with HCC or malignancy: size stability > 2 years, size reduction, and undistorted vessels. Like all the AFs favoring benignity, these AFs had low positivity rates (8.7%, 0.9% and 1.6%, respectively). In addition, size stability and size reduction were sparsely reported (only 252 and 348 observations, respectively). This is likely because, like TG, these AFs require a prior imaging examination for comparison.

Results of this IPD meta-analysis highlight important considerations regarding the application of AFs in LI-RADS. First, our study shows strong associations between most AFs and HCC, malignancy, or benignity when adjusting for major features. The application of AFs is optional according to LI-RADS. However, as shown previously (28), our results suggest that mandatory incorporation of some AFs could improve the diagnostic performance of LI-RADS v2018. Second, although LI-RADS v2018 weights each AF equally, our study found differences in the strength of associations between each AF and HCC/malignancy (Figure 3.3). For example, there is a 6.5-fold difference in the odds of malignancy between transitional phase hypointensity (OR:2.3) and restricted diffusion (OR:14.5). Third, LI-RADS v2018 specifies that the LI-RADS category should not be adjusted when an observation exhibits both AFs favoring malignancy and benignity (29). A weighted system, based on the number of AFs present and their association with HCC, may be more appropriate, however we did not assess combinations of AFs. Fourth, LI-RADS v2018 does not allow AFs to up-categorize an observation from LR-4 to LR-5, to preserve near-perfect positive predictive value of LR-5 for HCC (29). However, when adjusting

for the presence of major features except TG, some AFs had stronger associations with HCC than some major features (Figure 3.3, Appendix 3.6 and Appendix 3.7). For example, nonenhancing capsule (OR:3.50) had a stronger association with HCC than size and enhancing capsule (Figure 3.3). Therefore, our results suggest that up-categorization of LR-4 observations based on the AFs with strong independent association with HCC may improve the sensitivity of LR-5 without sacrificing specificity. However, further studies are needed to investigate this approach. Finally, several AFs showed no significant association according to their intended purpose, and this is likely in part because of low positivity rates, low sample size, and/or non-applicability. Given their low prevalence and lack of association, the overall impact of these AFs on the diagnostic performance of LI-RADS is likely low.

Our study had several limitations. Compared to major features, AFs have not been as comprehensively evaluated in research studies assessing the diagnostic performance of LI-RADS, resulting in small sample sizes and/or low positivity rates for some AFs. Because the sample size for each AF differed, developing a multivariable model that included all AFs was not possible. In addition, few studies have reported data on TG, likely because TG requires a prior imaging examination for comparison. This limited the utility of the secondary analysis. Most studies contributing data to the LI-RADS IPD had a large proportion of observations in the LI-RADS 3-5 range, and there were low frequencies of AFs intended to suggest benignity. LI-RADS 2-3 observations and AFs suggestive of benignity may have been selectively avoided because several studies relied on a pathology-based reference standard (30). Our meta-analysis included observations categorized as LR-M, however, according to LI-RADS v2018, AFs do not have a role in categorizing observations as LR-M. Several studies were at risk of bias, including 12/22 studies in the patient and observation selection domain and 13/22 studies in the flow and timing domain. Three studies did not evaluate AFs for all observations in their study (31-33), increasing the risk of selection bias. Studies included in most analyses were highly heterogenous (I^2>75%), therefore additional factors in the sample population or study design may have influenced our results.

The strengths of this study include the IPD meta-analysis approach, which enabled data validation at the individual participant level, and not in aggregate as with traditional meta-analyses. The database provided a sufficiently large sample size to evaluate the independent

association of each AF with malignancy in general, HCC specifically, and benignity. This is the largest study to date evaluating AFs and these associations and included over 3000 observations from 20 studies.

In conclusion, this IPD meta-analysis found that most AFs are independently associated with HCC, malignancy, and benignity as intended by LI-RADS, when adjusting for major features. These associations were determined by excluding TG as a major feature, which enabled larger sample sizes. Some AFs did not show associations according to their intended use, likely because of low positivity rates and, in some cases, low sample sizes. Our results highlight areas for refining LI-RADS. Future revisions of LI-RADS would be better informed with standardized study designs that reduce the risk of bias and consistently report all major and ancillary features.

Tables and Figures

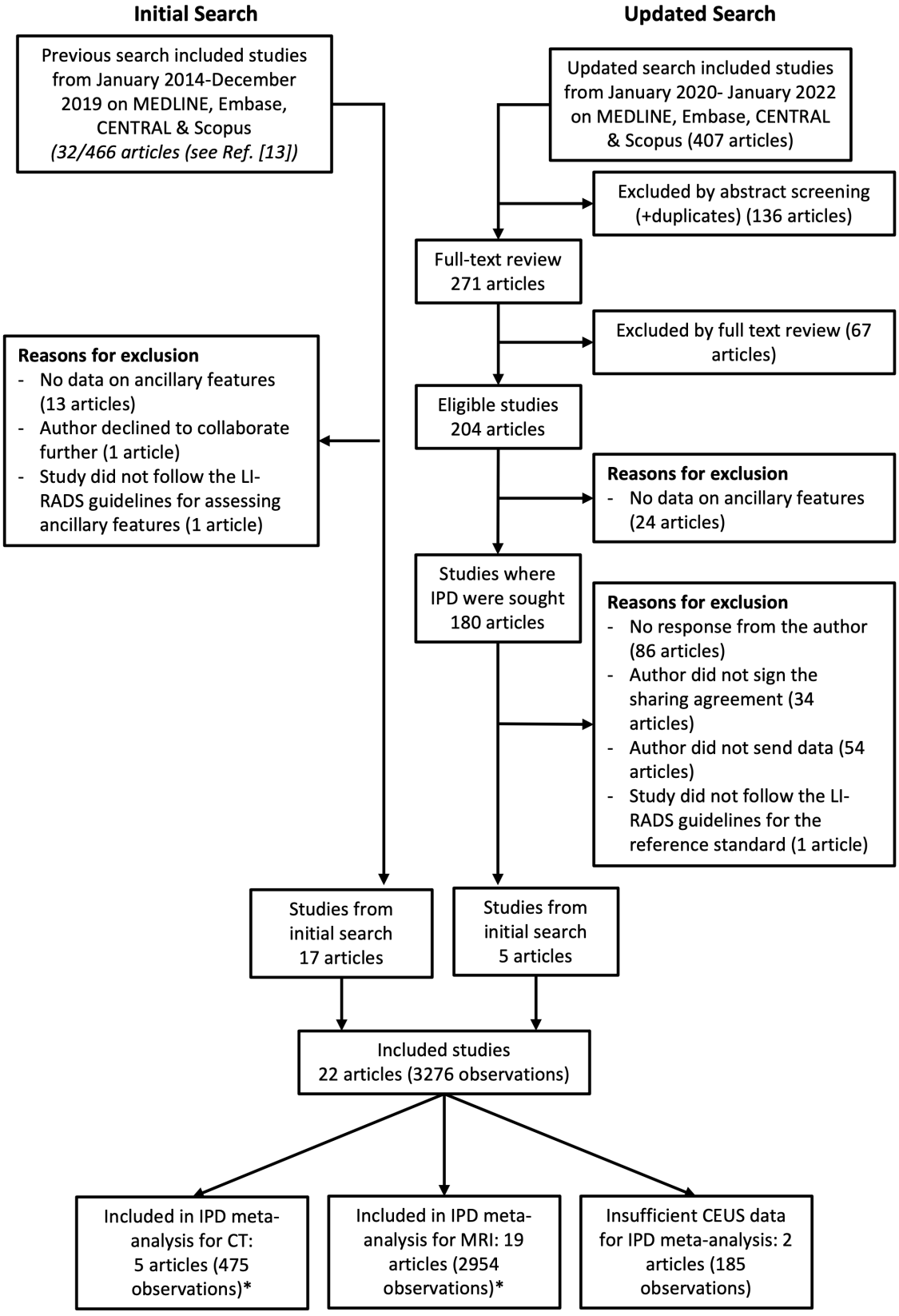


Figure 3.1. Flow diagram of study inclusion

*Four studies evaluated AFs using CT and MRI (22-25).

IPD: Individual patient data; CT: computed tomography; MRI: magnetic resonance imaging; CEUS: contrast enhanced ultrasound.

Table 3.1. Characteristics of included studies*

Study details					Imaging technique				Observation data												Reference standard
Study	Journal	Country	Design	Prevailing risk factor	Imaging modality	IV contrast agent type	LI-RADS version	No. readers	No. liver observations / No. patients	Median size, mm (IQR)	No. of HCC (% of total)	No. overall malignancy (% of total)	No. benign (% of total)	No. LR-1	No. LR-2	No. LR-3	No. LR-4	No. LR-5	No. LR-TIV/5V	No. LR-M	
Alhasan A 2019 (34)	Abdom Radiol (NY)	Canada	Retrospective cohort	Cirrhosis >> HBV	CT	ECA	v2017	2	102/59	22 (15-40)	72 (71)	76 (75)	26 (25)	0	4	15	29	41	10	3	Pathology
Allen BC 2018 (31)	AJR Am J Roentgenol	USA	Retrospective case-control	Cirrhosis >> HBV	MRI	HPB	v2014	3	16/16	29.5 (18.5-64.5)	13 (81)	15 (94)	1 (6)	0	0	3	3	10	0	0	Pathology and CCRS
Cerny M 2018 (28)	Radiology	Canada	Retrospective cohort	Cirrhosis >> HBV	MRI	ECA	v2014	2	195/100	17 (10-24)	110 (56)	119 (61)	76 (39)	1	12	55	53	57	2	15	Pathology and CCRS
Chen J 2019 (35)**	AJR Am J Roentgenol	China	Retrospective cohort	HBV > cirrhosis	MRI	ECA	v2018	2	104/104	47.5 (10-24)	104 (100)	104 (100)	0 (0)	0	0	0	0	104	0	0	Pathology
Choi JY 2022 (36)	Abdom Radiol (NY)	South Korea	Retrospective cohort	HBV >> Cirrhosis	MRI	HBP	v2018	2	279/253	26 (18-35)	247 (89)	268 (96)	11 (4)	0	0	18	46	190	6	19	Pathology
Fraum TJ 2018 (23)	Radiology	USA	Retrospective cohort	Cirrhosis >> HBV	CT, MRI	HPB	v2014	2	CT: 7/7 MRI: 213/213 Total: 220/220	CT: 104 (38-132) MRI: 31 (22-50)	CT: 4 (57) MRI: 132 (62) Total: 136 (62)	CT: 7 (100) MRI: 171 (80) Total: 178 (81)	CT: 0 (0) MRI: 42 (20) Total: 42 (19)	CT: 0 MRI: 3 Total: 3	CT: 0 MRI: 14 Total: 14	CT: 0 MRI: 8 Total: 8	CT: 0 MRI: 32 Total: 32	CT: 2 MRI: 96 Total: 98	CT: 2 MRI: 18 Total: 20	CT: 3 MRI: 42 Total: 45	Pathology
Fraum TJ 2020 (24)	European Radiology	USA	Retrospective cohort	Cirrhosis >> HBV	CT, MRI	NR	v2018	2	CT: 28/28 MRI: 92/92 Total: 120/120	CT: 4 (3-5.4) MRI: 3.5 (2.3-5.7)	CT: 15 (54) MRI: 41 (45) Total: 56 (47)	CT: 28 (100) MRI: 92 (100) Total: 120 (100)	CT: 0 (0) MRI: 0 (0) Total: 0 (0)	CT: 0 MRI: 0 Total: 0	CT: 0 MRI: 0 Total: 0	CT: 0 MRI: 2 Total: 2	CT: 1 MRI: 10 Total: 11	CT: 13 MRI: 27 Total: 40	CT: 0 MRI: 13 Total: 13	CT: 14 MRI: 40 Total: 54	Pathology
Hu J 2020 (37)	Abdom Radiol (NY)	Canada	Prospective cohort	Cirrhosis >> HBV	CEUS	Blood pool	v2017	2	25/24	17 (13-25)	16 (64)	17 (68)	8 (32)	0	3	5	1	16	0	0	Pathology and CCRS
Jiang H 2019 (38)	Cancer Imaging	China	Prospective cohort	HBV >> cirrhosis	MRI	HPB	v2018	2	272/272	50.7 (29.7-75)	215 (79)	254 (93)	18 (7)	1	3	4	28	151	57	28	Pathology and CCRS
Jeon SK 2018 (39)	European Radiology	South Korea	Retrospective case-control	HBV >> Cirrhosis	MRI	HBP	v2017	2	140/140	32 (22-49.5)	70 (50)	140 (100)	0 (0)	0	0	0	21	67	2	50	Pathology
Kang Z 2019 (40)	Eur J Radiol Open	China	Retrospective cohort	Cirrhosis, no other details	MRI	ECA	v2014	2	19/19	27 (16-40)	15 (79)	17 (89)	2 (11)	0	0	4	2	11	1	1	Pathology and CCRS
Kierans AS 2019 (41)	J Magn Reson Imaging	USA	Retrospective cohort	Cirrhosis > HBV	MRI	ECA, HPB	v2017	3	144/114	37.4 (10-37)	82 (57)	90 (63)	54 (38)	5	8	45	25	41	10	10	Pathology and CCRS
Kim DH 2019 (27)	J Hepatol	South Korea	Retrospective cohort	HBV > cirrhosis	MRI	HPB	v2018	NR	372 / 258	18 (11-25)	273 (73)	291 (78)	81 (22)	0	0	18	154	180	4	16	Pathology and CCRS
Kim DH 2021 (42)	J Magn Reson Imaging	South Korea	Retrospective cohort	HBV >> Cirrhosis	MRI	HBP	v2018	3	113/113	34 (25-50)	0 (0)	113 (100)	0 (0)	0	0	0	24	32	13	44	Pathology
Lee CM 2021 (25)	Hepatology International	South Korea	Retrospective cohort	HBV >> Cirrhosis	CT, MRI	HBP	v2018	2	CT: 268/218 MRI: 291/222 Total: 291/222	CT: 20 (13-27) MRI: 18 (12-27)	CT:206 (77) MRI: 208 (71) Total: 208 (71)	CT: 218 (81) MRI: 220 (76) MRI: 220 (76)	CT: 50 (19) MRI: 71 (24) Total: 71 (24)	CT: 0 MRI: 0	CT: 0 MRI: 0	CT: 72 MRI: 15	CT: 37 MRI: 104	CT: 152 MRI: 154	CT: 3 MRI: 2	CT: 4 MRI: 16	Pathology and CCRS
Lim K 2021 (43)	BR J Radiol	South Korea	Retrospective cohort	Cirrhosis, no other details	MRI	HBP	v2018	2	161/112	11 (8-16)	107 (66)	107 (66)	54 (34)	0	0	15	146	0	0	0	Pathology and CCRS
Makoyeva A 2020 (44)	Radiol Imaging Cancer	USA	Retrospective cohort	Cirrhosis >> HBV	CEUS	Lipid microspheres	v2016	3	160/160	24 (16-35)	126 (79)	128 (80)	32 (20)	4	1	24	8	114	7	2	Pathology and CCRS
Ronot M 2018 (26)	J Hepatol	France	Prospective cohort	Cirrhosis >> HBV	CT, MRI	ECA	v2014	1	CT: 70/62 MRI: 93/76 Total: 93/76	CT: 15 (12-19) MRI: 15 (12-18)	CT: 20 (29) MRI: 23 (25) Total: 23 (25)	CT: 20 (29) MRI: 23 (25) Total: 23 (25)	CT: 0 (0) MRI: 0 (0) Total: 0 (0)	CT: 0 MRI: 0	CT: 0 MRI: 0	CT: 43 MRI: 93	CT: 10 MRI: 0	CT: 17 MRI: 0	CT: 0 MRI: 0	CT: 0 MRI: 0	Pathology
Rosiak G 2018 (45)	Biomed Res Int	Poland	Retrospective cohort	Cirrhosis >> HBV	MRI	HPB	v2017	2	69/32	21 (12-35)	50 (72)	50 (72)	19 (28)	0	0	18	13	38	0	0	Pathology
Song JS 2019 (33)	Eur Radiol	South Korea	Retrospective cohort	HBV ~ cirrhosis	MRI	HPB	v2014	2	77/64	1.7 (1.3-2.3)	77 (100)	77 (100)	0 (0)	0	0	ECA: 1 HPB: 1	ECA: 25 HPB: 39	ECA: 51 HPB: 37	0	0	Pathology and CCRS
van der Pol CB 2021 (46)	AJR Am J Roentgenol	Canada	Retrospective cohort	Cirrhosis >> HBV	MRI	ECA, HPB	v2018	2	222/81	11 (8-17)	72 (32)	72 (32)	150 (68)	23	33	68	42	56	0	0	Pathology and CCRS
Zhang L 2019 (47)	Front. Oncol.	China	Retrospective cohort	HBV >> Cirrhosis	MRI	ECA	v2018	2	82/81	22.5 (17-27)	82 (100)	82 (100)	0 (0)	0	0	7	7	68	0	0	Pathology

HBV: hepatitis B virus; MRI: magnetic resonance imaging; CT: computed tomography; CEUS: contrast enhanced ultrasound; ECA: extracellular contrast agent; HPB: hepatobiliary contrast agent; NR: not reported; CCRS: composite reference standard.

*Values may not match published data as participants were excluded on an observation level when applicable.

**Chen J 2019 (34) had an overlapping study population with Zhang L 2019 (46). Forty-five duplicate observations were removed from the study by Chen J et al.

Table 3.2. Observation diagnosis and reference standard Observations evaluated on CT/MRI and CEUS were incorporated into this table.*

Final Diagnosis	**Total No. of observations**	**Confirmed with Histology**	**Confirmed with Composite Reference Standard** †
HCC	2154	1752	402
Non-HCC malignancy	407	405	2
Benign	645	248	397

*Seventy observations from Ronot et al. (25) were excluded from the table because they were diagnosed as non-HCC.

† The reference standard is available as supplementary appendix 3.1

Abbreviations- HCC: hepatocellular carcinoma

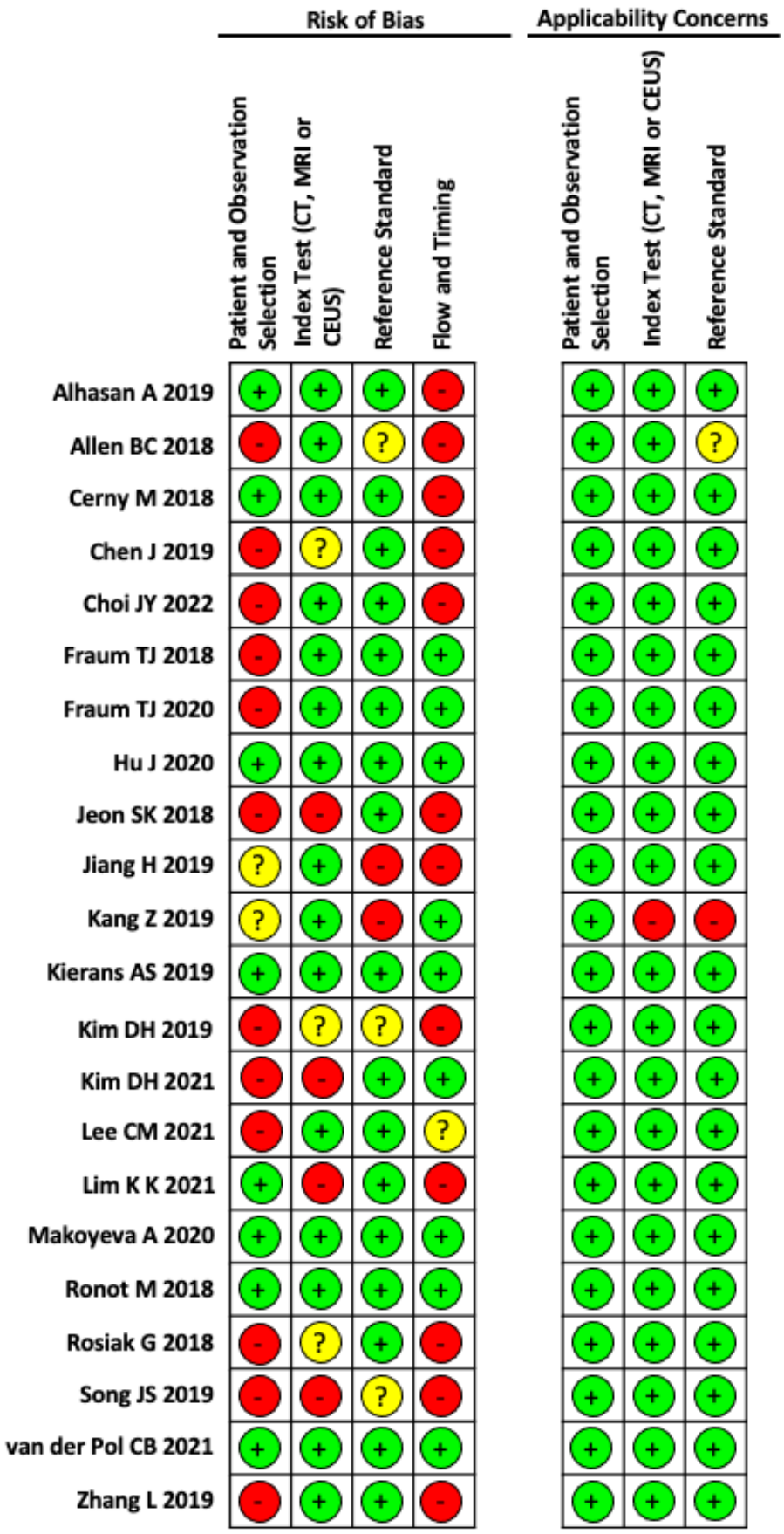


Figure 3.2. Risk of bias and applicability concerns for each study.

Table 3.3. Multivariable analysis of each ancillary feature, adjusted for major features (arterial phase hyperenhancement (APHE), size, nonperipheral washout and enhancing capsule) to determine the odds of Hepatocellular Carcinoma (HCC) and malignancy.

Feature	Association with (Based on LIRADS v2018)	Study sources	HCC				Malignancy			
			Observations (Patients, Studies)	HCC observations n(%)	Positive Data points n(%)	OR (95%CI)	Observations (Patients, Studies)	Malignant observations n(%)	Positive Data points n(%)	OR (95%CI)
US visibility as discrete nodule	Malignancy	(33), (24), (26), (22), (39), (25)**	904 (676,6)	606 (67.0)	798 (88.3)	**3.18 (1.77-5.73)***	811 (600,5)	628 (77.4)	739 (91.1)	2.09 (0.83-5.26)
Subthreshold growth	Malignancy	(33), (26), (22), (39)	267 (207,4)	186 (69.7)	80 (30.0)	**3.21 (1.26-8.17)***	267 (207,4)	203 (76.0)	80 (30.0)	**5.03 (1.64-15.46)***
Restricted diffusion	Malignancy	(27), (34), (41), (24), (35), (26), (22), (23), (42), (37), (38), (39), (40), (25)**, (44), (45), (46)	2817 (2262,17)	1806 (64.1)	1753 (62.2)	**3.18 (2.52-4.02)***	2724 (2186,16)	2162 (79.4)	1723 (63.3)	**14.45 (9.82-21.27)***
Mild-moderate T2 hyperintensity	Malignancy	(27), (34), (41), (24), (35), (26), (22), (23), (42), (37), (38), (39), (40), (25)**, (44), (32), (45), (46)	2938 (2346,18)	1908 (64.9)	2054 (69.9)	**3.05 (2.39-3.89)***	2845 (2270,17)	2267 (79.7)	2012 (70.7)	**10.18 (7.17-14.44)***
Corona enhancement	Malignancy	(33), (27), (34), (41), (24), (35), (26), (22), (23), (42), (37), (39), (40), (25)**, (44), (32), (45), (46)	2935 (2300,18)	1929 (65.7)	246 (8.38)	1.18 (0.75-1.87)	2842 (2224,17)	2238 (78.9)	241 (8.5)	**4.89 (1.71-14.01)***
Fat sparing in solid mass	Malignancy	(33), (27), (34), (41), (24), (35), (26), (22), (23), (42), (37), (39), (40), (25)**, (44), (32), (45), (46)	2771 (2213,18)	1833 (66.2)	109 (3.9)	0.81 (0.44-1.49)	2678 (2137,17)	2132 (79.6)	97 (3.6)	0.79 (0.30-2.09)
Iron sparing in solid mass	Malignancy	(27), (34), (41), (24), (35), (26), (22), (23), (37), (39), (40), (25)**, (44), (32), (45)	2464 (2006,16)	1634 (66.3)	159 (6.5)	**2.63 (1.30-5.31)***	2371 (1930,15)	1918 (80.9)	151 (6.4)	2.69 (0.52-13.92)
Transitional phase hypointensity	Malignancy	(41), (24), (35), (26), (42), (37), (38), (40), (44), (32), (45)	1871 (1516,11)	1272 (68.0)	1486 (79.4)	**2.08 (1.48-2.93)***	1871 (1516,11)	1547 (82.7)	1486 (79.4)	**2.28 (1.45-3.59)***

Hepatobiliary phase hypointensity	Malignancy	(41), (24), (35), (26), (42), (37), (38), (40), (44), (32), (45)	1914 (1533,11)	1307 (68.3)	1682 (87.9)	**2.06 (1.42-3.00)***	1914 (1533,11)	1578 (82.5)	1682 (87.9)	**3.15 (1.95-5.07)***
Growth (threshold + subthreshold)	Malignancy	(33), (27), (22), (23), (39)	395 (317,5)	253 (64.1)	143 (36.2)	**2.68 (1.42-5.07)***	395 (317,5)	320 (81.0)	143 (36.2)	**4.36 (1.87-10.17)***
Nonenhancing "capsule"	HCC	(33), (34), (41), (24), (35), (26), (23), (42), (37), (39), (40), (44), (32), (45), (46)	2426 (1903,15)	1659 (68.4)	94 (3.9)	**3.50 (1.53-8.01)***	2426 (1903,15)	1940 (80.0)	94 (3.9)	2.06 (0.61-6.91)
Nodule-in-nodule	HCC	(33), (27), (34), (41), (24), (35), (26), (22), (23), (42), (37), (38), (39), (40), (44), (32), (45), (46)	2982 (2364,18)	1976 (66.3)	187 (6.3)	**2.15 (1.15-4.04)***	2982 (2982,18)	2378 (79.8)	187 (6.3)	1.43 (0.46-4.43)
Mosaic architecture	HCC	(33), (27), (34), (41), (24), (35), (26), (22), (23), (42), (37), (39), (40), (25)**, (44), (32), (45), (46)	2935 (2300,18)	1929 (65.7)	453 (15.4)	**2.53 (1.58-4.04)***	2842 (2224,17)	2238 (78.9)	447 (15.7)	**3.26 (1.25-8.51)***
Blood products in mass	HCC	(33), (27), (34), (41), (24), (35), (26), (22), (23), (37), (38), (39), (40), (25)**, (44), (32), (45), (46)	2868 (2297,18)	1860 (64.9)	263 (9.2)	1.28 (0.83-1.96)	2775 (2221,17)	2234 (80.5)	260 (9.4)	2.31 (0.81-6.57)
Fat in mass, more than adjacent liver	HCC	(33), (27), (34), (41), (24), (35), (26), (22), (23), (42), (37), (38), (39), (40), (25)**, (44), (32), (45), (46)	3063 (2429,19)	1990 (65.0)	437 (14.3)	**1.44 (1.03-2.01)***	2970 (2353,18)	2366 (79.7)	431 (14.5)	0.71 (0.42-1.19)
Size stability > 2 years	Benignity	(33), (27), (22), (39)	252 (202,4)	185 (73.4)	22 (8.7)	0.45 (0.13-1.55)	252 (202,4)	204 (81.0)	22 (8.7)	0.63 (0.18-2.14)
Size reduction	Benignity	(33), (27), (22), (39)	348 (244,4)	230 (66.1)	3 (0.9)	0.28 (0.01-12.34)	348 (244,4)	252 (72.4)	3 (0.9)	0.17 (0.003-10.75)
Parallels blood pool	Benignity	(33), (27), (24), (35), (26), (22), (23), (37), (39), (40), (44), (32), (45), (46)	2464 (1895,14)	1695 (68.8)	24 (1.0)	**0.07 (0.01-0.49)***	2464 (1895,14)	1914 (77.7)	24 (1.0)	**0.02 (0.002-0.24)***
Undistorted vessels	Benignity	(33), (27), (24), (35), (26), (22), (23), (37), (39), (40), (44), (32), (45), (46)	2464 (1895,14)	1695 (68.8)	39 (1.6)	0.90 (0.30-2.68)	2464 (1895,14)	1914 (77.7)	39 (1.6)	0.72 (0.17-3.12)

Iron in mass, more than liver	Benignity	(24), (35), (26), (22), (23), (37), (39), (40), (44), (45), (46)	2083 (1665,11)	1432 (68.8)	31 (1.5)	0.27 (0.07-1.11)	2083 (1665,11)	1635 (78.5)	31 (1.5)	**0.07 (0.01-0.44)***
Marked T2 hyperintensity	Benignity	(27), (24), (35), (26), (22), (23), (42), (37), (39), (40), (44), (32), (45), (46)	2488 (1913,14)	1711 (68.8)	78 (3.1)	**0.18 (0.07-0.45)***	2488 (1913,14)	1910 (76.8)	78 (3.1)	**0.06 (0.02-0.19)***
Hepatobiliary phase iso-intensity	Benignity	(30)***, (41), (24), (35), (26), (42), (37), (40), (44), (32), (45)	1802 (1408,11)	1253 (69.5)	110 (6.1)	**0.46 (0.27-0.80)***	1786 (1402,10)	1445 (80.9)	109 (6.1)	**0.16 (0.08-0.30)***

*Significant association ($p<0.05$)

**Ronot study was used to assess HCC only

***Allen study was used to assess HCC only

Abbreviations- HCC: hepatocellular carcinoma; US: ultrasound; CI: confidence interval

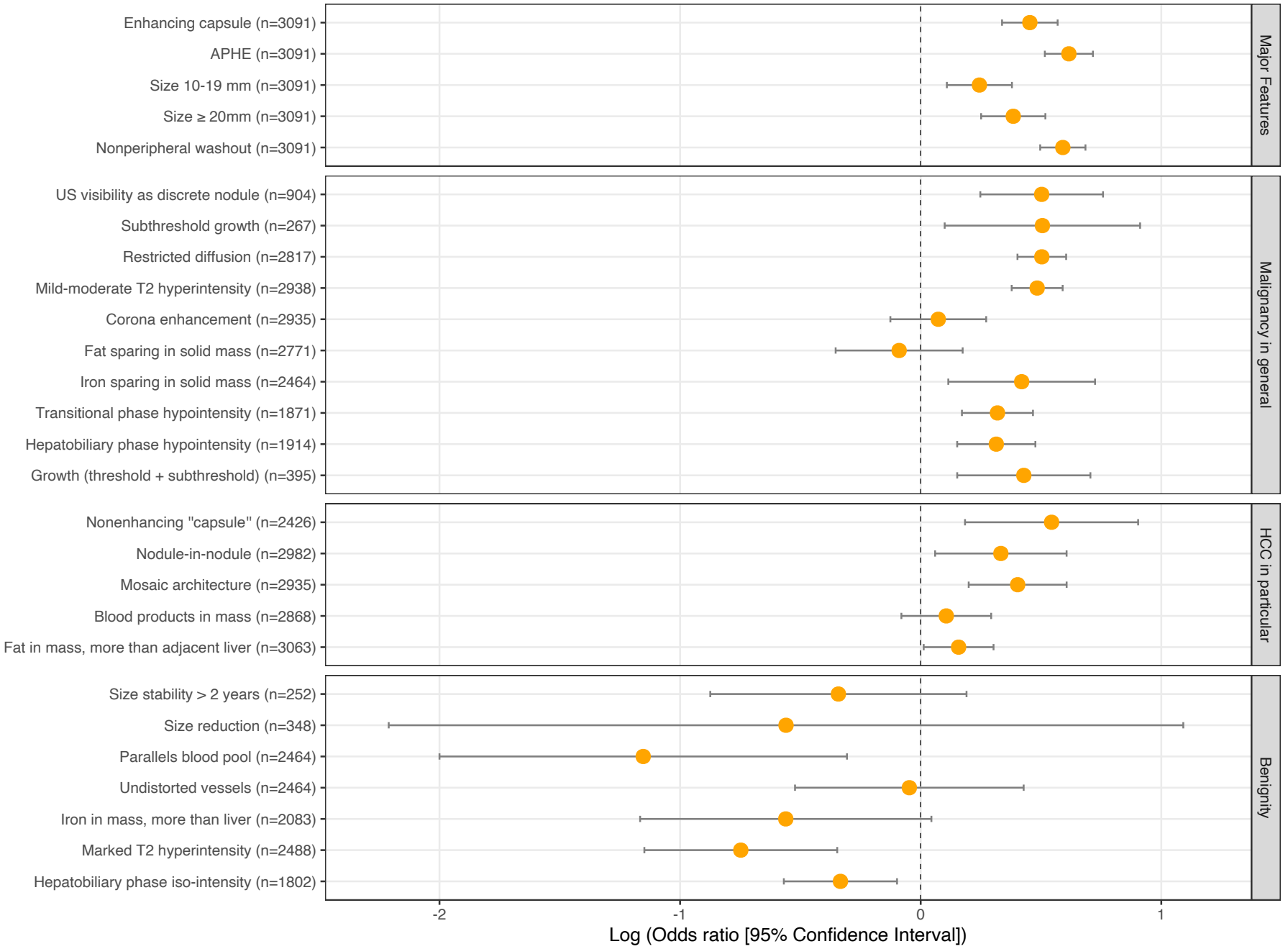

Enhancing capsule (n=3091)
APHE (n=3091)
Size 10-19 mm (n=3091)
Size ≥ 20mm (n=3091)
Nonperipheral washout (n=3091)
Major Features
US visibility as discrete nodule (n=904)
Subthreshold growth (n=267)
Restricted diffusion (n=2817)
Mild-moderate T2 hyperintensity (n=2938)
Corona enhancement (n=2935)
Fat sparing in solid mass (n=2771)
Iron sparing in solid mass (n=2464)
Transitional phase hypointensity (n=1871)
Hepatobiliary phase hypointensity (n=1914)
Growth (threshold + subthreshold) (n=395)
Malignancy in general
Nonenhancing "capsule" (n=2426)
Nodule-in-nodule (n=2982)
Mosaic architecture (n=2935)
Blood products in mass (n=2868)
Fat in mass, more than adjacent liver (n=3063)
HCC in particular
Size stability > 2 years (n=252)
Size reduction (n=348)
Parallels blood pool (n=2464)
Undistorted vessels (n=2464)
Iron in mass, more than liver (n=2083)
Marked T2 hyperintensity (n=2488)
Hepatobiliary phase iso-intensity (n=1802)
Benignity
-2
-1
0
1
Log (Odds ratio [95% Confidence Interval])

Figure 3.3. Log odds ratios and 95% confidence intervals of each major feature and ancillary feature, adjusted for four major features (APHE, size, nonperipheral washout and enhancing capsule) to determine the odds of hepatocellular carcinoma (HCC) in particular. AFs were divided by their association to malignancy, ±HCC or benignity according to LI-RADS v2018 guidelines (axis on the right). Due to varying sample sizes, the odds ratios for each AF were determined using different models. Odds ratios for the major features were derived using a single model without any ancillary features. (APHE: arterial phase hyperenhancement; HCC: hepatocellular carcinoma; US: ultrasound

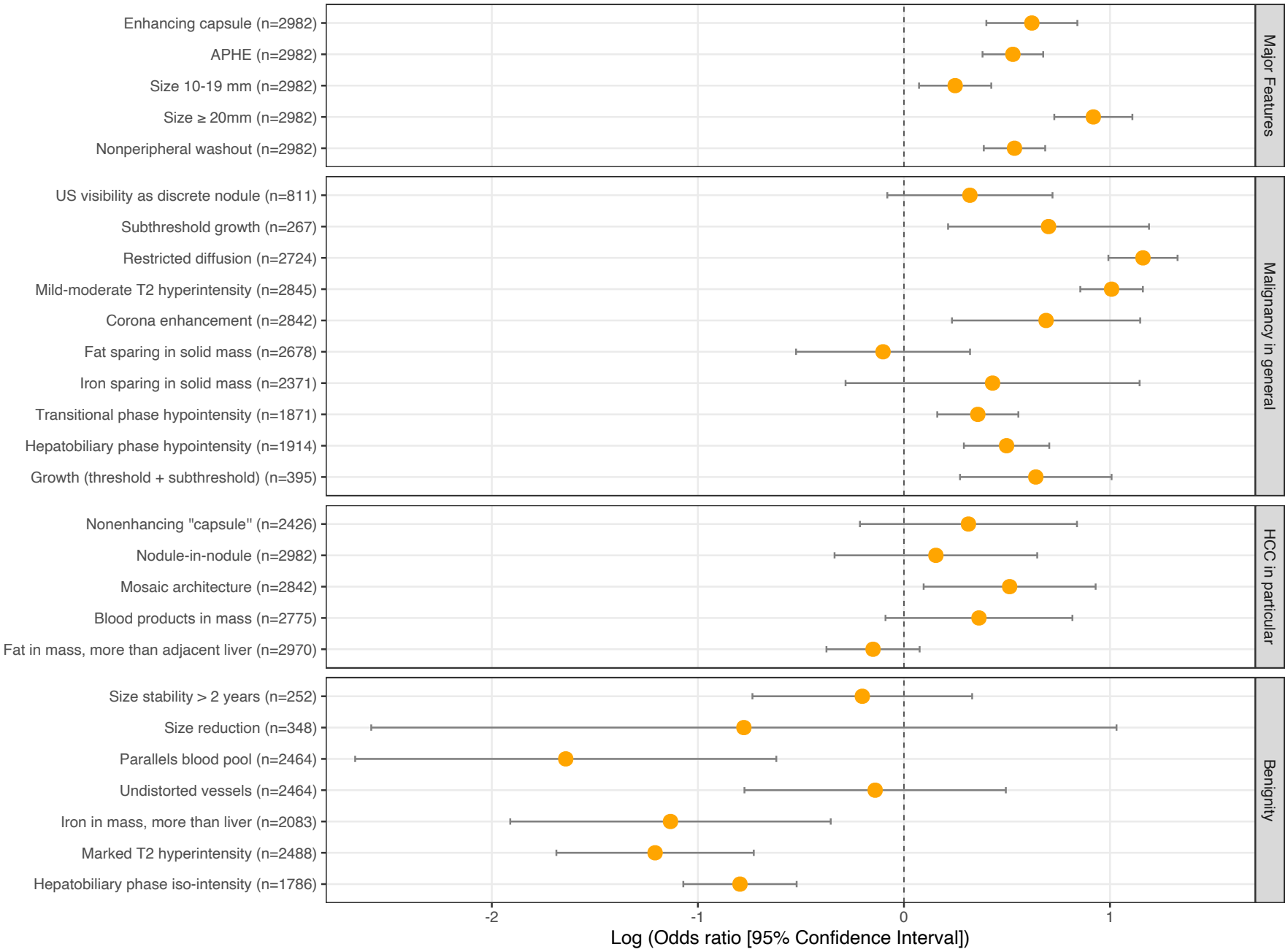

Enhancing capsule (n=2982)
APHE (n=2982)
Size 10-19 mm (n=2982)
Size ≥ 20mm (n=2982)
Nonperipheral washout (n=2982)
Major Features
US visibility as discrete nodule (n=811)
Subthreshold growth (n=267)
Restricted diffusion (n=2724)
Mild-moderate T2 hyperintensity (n=2845)
Corona enhancement (n=2842)
Fat sparing in solid mass (n=2678)
Iron sparing in solid mass (n=2371)
Transitional phase hypointensity (n=1871)
Hepatobiliary phase hypointensity (n=1914)
Growth (threshold + subthreshold) (n=395)
Malignancy in general
Nonenhancing "capsule" (n=2426)
Nodule-in-nodule (n=2982)
Mosaic architecture (n=2842)
Blood products in mass (n=2775)
Fat in mass, more than adjacent liver (n=2970)
HCC in particular
Size stability > 2 years (n=252)
Size reduction (n=348)
Parallels blood pool (n=2464)
Undistorted vessels (n=2464)
Iron in mass, more than liver (n=2083)
Marked T2 hyperintensity (n=2488)
Hepatobiliary phase iso-intensity (n=1786)
Benignity
-2
-1
0
1
Log (Odds ratio [95% Confidence Interval])

Figure 3.4. Log odds ratios and 95% confidence intervals of each major feature and ancillary feature, adjusted for four major features (APHE, size, nonperipheral washout and enhancing capsule) to determine the odds of malignancy in general. AFs were divided by their association to malignancy, hepatocellular carcinoma (HCC) or benignity according to LI-RADS v2018 guidelines (axis on the right). Due to varying sample sizes, the odds ratios for each AF were determined using different models. Odds ratios for the major features were derived using a single model without any ancillary features. (APHE: arterial phase hyperenhancement; HCC: hepatocellular carcinoma; US: ultrasound)

Table 3.4. The within study (I^2) and between study-variance (tau^2) in each random effects model, after adjusting for four major features (arterial phase hyperenhancement, size, nonperipheral washout and enhancing capsule)

Ancillary Feature	**HCC**		**Malignancy**	
	I^2 [%(95%CI)]	**Tau^2**	**I^2 [%(95%CI)]**	**Tau^2**
US visibility as discrete nodule	92.7 (86.8-96.0)	0.574	0.0 (0.0-79.2)	0
Subthreshold growth	0.0 (0.0-84.7)	0	15.8 (0.0-87.1)	0
Restricted diffusion	94.2 (92.1-95.8)	5.025	94.5 (92.5-96.0)	6.351
Mild-moderate T2 hyperintensity	93.9 (91.7-95.5)	6.571	94.2 (92.1-95.8)	7.371
Corona enhancement	93.7 (91.4-95.4)	6.469	94.3 (92.2-95.8)	5.597
Fat sparing in solid mass	93.6 (91.2-95.3)	6.463	94.0 (91.9-95.7)	5.601
Iron sparing in solid mass	94.3 (92.1-95.8)	8.081	94.7 (92.7-96.2)	6.125
Transitional phase hypointensity	93.4 (90.1-95.6)	5.962	94.6 (92.1-96.3)	5.658
Hepatobiliary phase hypointensity	93.5 (90.2-95.6)	5.686	94.6 (92.1-96.3)	5.809
Growth (threshold + subthreshold)	79.5 (51.4-91.3)	0.184	0.8 (0.0-79.4)	1.887
Nonenhancing "capsule"	93.3 (90.5-95.3)	8.373	94.6 (92.5-96.1)	6.662
Nodule-in-nodule	92.7 (89.9-94.7)	6.114	93.9 (91.7-95.5)	6.860
Mosaic architecture	93.7 (91.4-95.4)	6.469	94.3 (92.2-95.8)	5.597
Blood products in mass	93.9 (91.7-95.5)	6.553	94.2 (92.0-95.7)	7.390
Fat in mass, more than adjacent liver	93.6 (91.3-95.2)	5.933	93.8 (91.6-95.5)	6.892
Size stability > 2 years	0.0 (0.0-84.7)	0	0.0 (0.0-84.7)	0
Size reduction	17.0 (0.0-87.3)	0	64.4 (0.0-87.9)	0.078
Parallels blood pool	94.1 (91.6-95.8)	2.056	95.3 (93.5-96.6)	3.659
Undistorted vessels	94.1 (91.6-95.8)	2.056	95.3 (93.5-96.6)	3.659
Iron in mass, more than liver	95.2 (93.1-96.7)	1.384	96.2 (94.7-97.3)	3.267
Marked T2 hyperintensity	94.0 (91.5-95.8)	2.093	95.3 (93.5-96.6)	3.602
Hepatobiliary phase iso-intensity	91.8 (87.4-94.7)	5.769	95.2 (92.8-96.7)	3.940

Abbreviations- US: ultrasound

References

1. American College of Radiology Liver Reporting and Data System (LI-RADS). www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/LI-RADS. Published 2021.9 Jan 2023.

2. Marrero JA, Kulik LM, Sirlin CB, Zhu AX, Finn RS, Abecassis MM, Roberts LR, Heimbach JK. Diagnosis, Staging, and Management of Hepatocellular Carcinoma: 2018 Practice Guidance by the American Association for the Study of Liver Diseases. Hepatology 2018;68(2):723-750. doi: 10.1002/hep.29913

3. Singal AG, Llovet JM, Yarchoan M, Mehta N, Heimbach JK, Dawson LA, Jou JH, Kulik LM, Agopian VG, Marrero JA, Mendiratta-Lala M, Brown DB, Rilling WS, Goyal L, Wei AC, Taddei TH. AASLD practice guidance on prevention, diagnosis, and treatment of hepatocellular carcinoma. Hepatology 9900:10.1097/HEP.0000000000000466. doi: 10.1097/hep.0000000000000466

4. Kambadakone AR, Fung A, Gupta RT, Hope TA, Fowler KJ, Lyshchik A, Ganesan K, Yaghmai V, Guimaraes AR, Sahani DV, Miller FH. LI-RADS technical requirements for CT, MRI, and contrast-enhanced ultrasound. Abdominal Radiology 2018;43(1):56-74. doi: 10.1007/s00261-017-1325-y

5. Corwin MT, Lee AY, Fananapazir G, Loehfelm TW, Sarkar S, Sirlin CB. Nonstandardized Terminology to Describe Focal Liver Lesions in Patients at Risk for Hepatocellular Carcinoma: Implications Regarding Clinical Communication. AJR Am J Roentgenol 2018;210(1):85-90. doi: 10.2214/AJR.17.18416

6. van der Pol CB, Lim CS, Sirlin CB, McGrath TA, Salameh JP, Bashir MR, Tang A, Singal AG, Costa AF, Fowler K, McInnes MDF. Accuracy of the Liver Imaging Reporting and Data System in Computed Tomography and Magnetic Resonance Image Analysis of Hepatocellular Carcinoma or Overall Malignancy-A Systematic Review. Gastroenterology 2019;156(4):976-986. doi: 10.1053/j.gastro.2018.11.020

7. Elsayes KM, Fowler KJ, Chernyak V, Elmohr MM, Kielar AZ, Hecht E, Bashir MR, Furlan A, Sirlin CB. User and system pitfalls in liver imaging with LI-RADS. J Magn Reson Imaging 2019;50(6):1673-1686. doi: 10.1002/jmri.26839

8. Bashir M, Kielar A, Zaki I, McInnes MDF, Costa AF, Fetzer DT, Hecht E, Kieran A, Heba E, Surabhi V, Yacoub J, Fowler K, Elsayes KM, Chernyak V, Sirlin CB. A Call to Action: Most

pressing research gaps identified by the LI-RADS Evidence-Based Working Group. 105th Scientific Assembly and Annual Meeting, Radiological Society of North America. Chicago IL2019.
9. Tang A, Bashir MR, Corwin MT, Cruite I, Dietrich CF, Do RKG, Ehman EC, Fowler KJ, Hussain HK, Jha RC, Karam AR, Mamidipalli A, Marks RM, Mitchell DG, Morgan TA, Ohliger MA, Shah A, Vu KN, Sirlin CB, Group L-REW. Evidence Supporting LI-RADS Major Features for CT- and MR Imaging-based Diagnosis of Hepatocellular Carcinoma: A Systematic Review. Radiology 2018;286(1):29-48. doi: 10.1148/radiol.2017170554
10. Shin J, Lee S, Yoon JK, Chung YE, Choi JY, Park MS. LI-RADS Major Features on MRI for Diagnosing Hepatocellular Carcinoma: A Systematic Review and Meta-Analysis. J Magn Reson Imaging 2021;54(2):518-525. doi: 10.1002/jmri.27570
11. Fowler KJ, Tang A, Santillan C, Bhargavan-Chatfield M, Heiken J, Jha RC, Weinreb J, Hussain H, Mitchell DG, Bashir MR, Costa EAC, Cunha GM, Coombs L, Wolfson T, Gamst AC, Brancatelli G, Yeh B, Sirlin CB. Interreader Reliability of LI-RADS Version 2014 Algorithm and Imaging Features for Diagnosis of Hepatocellular Carcinoma: A Large International Multireader Study. Radiology 2018;286(1):173-185. doi: 10.1148/radiol.2017170376
12. Kang JH, Choi SH, Lee JS, Park SH, Kim KW, Kim SY, Lee SS, Byun JH. Interreader Agreement of Liver Imaging Reporting and Data System on MRI: A Systematic Review and Meta-Analysis. J Magn Reson Imaging 2020;52(3):795-804. doi: 10.1002/jmri.27065
13. van der Pol CB, McInnes MDF, Salameh JP, Levis B, Chernyak V, Sirlin CB, Bashir MR, Allen BC, Burke LMB, Choi JY, Choi SH, Forner A, Fraum TJ, Giamperoli A, Jiang H, Joo I, Kang Z, Kierans AS, Kang HJ, Khatri G, Kim JH, Kim MJ, Kim SY, Kim YY, Kwon H, Lee JM, Lewis SC, McGinty KA, Mulazzani L, Park MS, Piscaglia F, Podgorska J, Reiner CS, Ronot M, Rosiak G, Song B, Song JS, Tang A, Terzi E, Wang J, Wang W, Wilson SR, Yokoo T. CT/MRI and CEUS LI-RADS Major Features Association with Hepatocellular Carcinoma: Individual Patient Data Meta-Analysis. Radiology 2022;302(2):326-335. doi: 10.1148/radiol.2021211244
14. Cerny M, Chernyak V, Olivie D, Billiard JS, Murphy-Lavallee J, Kielar AZ, Elsayes KM, Bourque L, Hooker JC, Sirlin CB, Tang A. LI-RADS Version 2018 Ancillary Features at MRI. Radiographics 2018;38(7):1973-2001. doi: 10.1148/rg.2018180052

15. McInnes MD, Bossuyt PM. Pitfalls of Systematic Reviews and Meta-Analyses in Imaging Research. Radiology 2015;277(1):13-21. doi: 10.1148/radiol.2015142779
16. Deeks JJ, Bossuyt P, Leeflang MM, Takwoingi Y, Cochrane Collaboration. Cochrane handbook for systematic reviews of diagnostic test accuracy. Hoboken, NJ London: Wiley-Blackwell ; The Cochrane Collaboration, 2023.
17. Cohen JF, Deeks JJ, Hooft L, Salameh JP, Korevaar DA, Gatsonis C, Hopewell S, Hunt HA, Hyde CJ, Leeflang MM, Macaskill P, McGrath TA, Moher D, Reitsma JB, Rutjes AWS, Takwoingi Y, Tonelli M, Whiting P, Willis BH, Thombs B, Bossuyt PM, McInnes MDF. Preferred reporting items for journal and conference abstracts of systematic reviews and meta-analyses of diagnostic test accuracy studies (PRISMA-DTA for Abstracts): checklist, explanation, and elaboration. Bmj 2021;372:n265. doi: 10.1136/bmj.n265
18. Salameh JP, Bossuyt PM, McGrath TA, Thombs BD, Hyde CJ, Macaskill P, Deeks JJ, Leeflang M, Korevaar DA, Whiting P, Takwoingi Y, Reitsma JB, Cohen JF, Frank RA, Hunt HA, Hooft L, Rutjes AWS, Willis BH, Gatsonis C, Levis B, Moher D, McInnes MDF. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. Bmj 2020;370:m2632. doi: 10.1136/bmj.m2632
19. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, McLeod L, Delacqua G, Delacqua F, Kirby J, Duda SN, Consortium RE. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019;95:103208. doi: 10.1016/j.jbi.2019.103208
20. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381. doi: 10.1016/j.jbi.2008.08.010
21. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM, Group Q-. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155(8):529-536. doi: 10.7326/0003-4819-155-8-201110180-00009
22. Doebler P. mada package for R. Vienna, Austria: R Foundation for Statistical Computing, 2015.

23. Fraum TJ, Tsai R, Rohe E, Ludwig DR, Salter A, Nalbantoglu I, Heiken JP, Fowler KJ. Differentiation of Hepatocellular Carcinoma from Other Hepatic Malignancies in Patients at Risk: Diagnostic Performance of the Liver Imaging Reporting and Data System Version 2014. Radiology 2018;286(1):158-172. doi: 10.1148/radiol.2017170114
24. Fraum TJ, Cannella R, Ludwig DR, Tsai R, Naeem M, LeBlanc M, Salter A, Tsung A, Shetty AS, Borhani AA, Furlan A, Fowler KJ. Assessment of primary liver carcinomas other than hepatocellular carcinoma (HCC) with LI-RADS v2018: comparison of the LI-RADS target population to patients without LI-RADS-defined HCC risk factors. Eur Radiol 2020;30(2):996-1007. doi: 10.1007/s00330-019-06448-6
25. Lee CM, Choi SH, Byun JH, Lee SJ, Kim SY, Won HJ, Shin YM, Kim PN. Combined computed tomography and magnetic resonance imaging improves diagnosis of hepatocellular carcinoma </= 3.0 cm. Hepatol Int 2021;15(3):676-684. doi: 10.1007/s12072-021-10190-x
26. Ronot M, Fouque O, Esvan M, Lebigot J, Aubé C, Vilgrain V. Comparison of the accuracy of AASLD and LI-RADS criteria for the non-invasive diagnosis of HCC smaller than 3 cm. J Hepatol 2018;68(4):715-723. doi: 10.1016/j.jhep.2017.12.014
27. Kim DH, Choi SH, Byun JH, Kang JH, Lim Y-S, Lee SJ, Kim SY, Won HJ, Shin YM, Kim P-N. Arterial subtraction images of gadoxetate-enhanced MRI improve diagnosis of early-stage hepatocellular carcinoma. Journal of Hepatology 2019;71(3):534-542. doi: https://doi.org/10.1016/j.jhep.2019.05.005
28. Cerny M, Bergeron C, Billiard J-S, Murphy-Lavallée J, Olivié D, Bérubé J, Fan B, Castel H, Turcotte S, Perreault P, Chagnon M, Tang A. LI-RADS for MR Imaging Diagnosis of Hepatocellular Carcinoma: Performance of Major and Ancillary Features. Radiology 2018;288(1):118-128. doi: 10.1148/radiol.2018171678
29. Chernyak V, Fowler KJ, Kamaya A, Kielar AZ, Elsayes KM, Bashir MR, Kono Y, Do RK, Mitchell DG, Singal AG, Tang A, Sirlin CB. Liver Imaging Reporting and Data System (LI-RADS) Version 2018: Imaging of Hepatocellular Carcinoma in At-Risk Patients. Radiology 2018;289(3):816-830. doi: 10.1148/radiol.2018181494
30. van der Pol CB, McInnes MDF, Salameh JP, Chernyak V, Tang A, Bashir MR, Group L-RI, Collaborators L-RIG. Impact of Reference Standard on CT, MRI, and Contrast-enhanced US LI-RADS Diagnosis of Hepatocellular Carcinoma: A Meta-Analysis. Radiology 2022;303(3):544-545. doi: 10.1148/radiol.212340

31. Allen BC, Ho LM, Jaffe TA, Miller CM, Mazurowski MA, Bashir MR. Comparison of Visualization Rates of LI-RADS Version 2014 Major Features With IV Gadobenate Dimeglumine or Gadoxetate Disodium in Patients at Risk for Hepatocellular Carcinoma. AJR Am J Roentgenol 2018;210(6):1266-1272. doi: 10.2214/AJR.17.18981
32. Kang HJ, Kim JH, Joo I, Han JK. Additional value of contrast-enhanced ultrasound (CEUS) on arterial phase non-hyperenhancement observations (≥ 2 cm) of CT/MRI for high-risk patients: focusing on the CT/MRI LI-RADS categories LR-3 and LR-4. Abdom Radiol (NY) 2020;45(1):55-63. doi: 10.1007/s00261-019-02132-x
33. Song JS, Choi EJ, Hwang SB, Hwang HP, Choi H. LI-RADS v2014 categorization of hepatocellular carcinoma: Intraindividual comparison between gadopentetate dimeglumine-enhanced MRI and gadoxetic acid-enhanced MRI. Eur Radiol 2019;29(1):401-410. doi: 10.1007/s00330-018-5559-z
34. Alhasan A, Cerny M, Olivie D, Billiard JS, Bergeron C, Brown K, Bodson-Clermont P, Castel H, Turcotte S, Perreault P, Tang A. LI-RADS for CT diagnosis of hepatocellular carcinoma: performance of major and ancillary features. Abdom Radiol (NY) 2019;44(2):517-528. doi: 10.1007/s00261-018-1762-2
35. Chen J, Zhou J, Kuang S, Zhang Y, Xie S, He B, Deng Y, Yang H, Shan Q, Wu J, Sirlin CB, Wang J. Liver Imaging Reporting and Data System Category 5: MRI Predictors of Microvascular Invasion and Recurrence After Hepatectomy for Hepatocellular Carcinoma. AJR Am J Roentgenol 2019;213(4):821-830. doi: 10.2214/AJR.19.21168
36. Choi JY, Ha J, Choi SH, Kang HJ, Kim SY, Kim KW. Comparison of gadoxetate disodium-enhanced MRI sequences for measuring hepatic observation size and its implication of LI-RADS classification. Abdom Radiol (NY) 2022;47(3):1024-1031. doi: 10.1007/s00261-021-03403-2
37. Hu J, Bhayana D, Burak KW, Wilson SR. Resolution of indeterminate MRI with CEUS in patients at high risk for hepatocellular carcinoma. Abdom Radiol (NY) 2020;45(1):123-133. doi: 10.1007/s00261-019-02181-2
38. Jiang H, Liu X, Chen J, Wei Y, Lee JM, Cao L, Wu Y, Duan T, Li X, Ma L, Song B. Man or machine? Prospective comparison of the version 2018 EASL, LI-RADS criteria and a radiomics model to diagnose hepatocellular carcinoma. Cancer Imaging 2019;19(1):84. doi: 10.1186/s40644-019-0266-9

39. Jeon SK, Joo I, Lee DH, Lee SM, Kang HJ, Lee KB, Lee JM. Combined hepatocellular cholangiocarcinoma: LI-RADS v2017 categorisation for differential diagnosis and prognostication on gadoxetic acid-enhanced MR imaging. Eur Radiol 2019;29(1):373-382. doi: 10.1007/s00330-018-5605-x
40. Kang Z, Wang N, Xu A, Wang L. Digital subtract angiography and lipiodol deposits following embolization in cirrhotic nodules of LIRADS category ≥3. Eur J Radiol Open 2019;6:106-112. doi: 10.1016/j.ejro.2019.02.005
41. Kierans AS, Makkar J, Guniganti P, Cornman-Homonoff J, Lee MJ, Pittman M, Askin G, Hecht EM. Validation of Liver Imaging Reporting and Data System 2017 (LI-RADS) Criteria for Imaging Diagnosis of Hepatocellular Carcinoma. J Magn Reson Imaging 2019;49(7):e205-e215. doi: 10.1002/jmri.26329
42. Kim DH, Choi SH, Kim DW, Lee SS, Lim YS, Kim SY, Kim HJ, Kim JH, Byun JH. Combined Hepatocellular-Cholangiocarcinoma: Magnetic Resonance Imaging Features and Prognosis According to Risk Factors for Hepatocellular Carcinoma. J Magn Reson Imaging 2021;53(6):1803-1812. doi: 10.1002/jmri.27528
43. Lim K, Kwon H, Cho J, Kim D, Kang E, Kim S. Added value of enhanced CT on LR-3 and LR-4 observation of Gd-EOB-DTPA MRI for the diagnosis of HCC: are CT and MR washout features interchangeable? Br J Radiol 2022;95(1132):20210738. doi: 10.1259/bjr.20210738
44. Makoyeva A, Kim TK, Jang HJ, Medellin A, Wilson SR. Use of CEUS LI-RADS for the Accurate Diagnosis of Nodules in Patients at Risk for Hepatocellular Carcinoma: A Validation Study. Radiol Imaging Cancer 2020;2(2):e190014. doi: 10.1148/rycan.2020190014
45. Rosiak G, Podgorska J, Rosiak E, Cieszanowski A. Comparison of LI-RADS v.2017 and ESGAR Guidelines Imaging Criteria in HCC Diagnosis Using MRI with Hepatobiliary Contrast Agents. Biomed Res Int 2018;2018:7465126. doi: 10.1155/2018/7465126
46. van der Pol CB, Dhindsa K, Shergill R, Zha N, Ferri M, Kagoma YK, Lee SY, Satkunasingham J, Wat J, Tsai S. MRI LI-RADS Version 2018: Impact of and Reduction in Ancillary Features. AJR Am J Roentgenol 2021;216(4):935-942. doi: 10.2214/ajr.20.23031
47. Zhang L, Kuang S, Chen J, Zhang Y, Zhao B, Peng H, Xiao Y, Fowler K, Wang J, Sirlin CB. The Role of Preoperative Dynamic Contrast-enhanced 3.0-T MR Imaging in Predicting Early Recurrence in Patients With Early-Stage Hepatocellular Carcinomas After Curative Resection. Front Oncol 2019;9:1336. doi: 10.3389/fonc.2019.01336

Chapter 4. Reflecting on the current landscape of diagnostic test accuracy studies in cancer-related imaging

List of authors

1. Haben Dawit, BSC. Clinical Epidemiology Program, Ottawa Hospital Research Institute. School of Epidemiology, Public Health and Preventative Medicine, Faculty of Medicine, University of Ottawa.

Preface

Manuscript Status

This is an unpublished manuscript.

Objective

The objective of this discussion is to summarize the findings of the individual chapters and connect it to the objective of this thesis.

Acknowledgements

I would like to thank Drs. Matthew McInnes, Rodney Breau and Brian Hutton for their support and feedback on my discussion.

Summary of Findings

The purpose of this thesis was to explore the role of non-invasive imaging in cancer diagnosis. Chapter 2 assessed the accuracy of biparamteric (bp)-MRI and multiparametric (mp)-MRI in diagnosing solid renal masses. Ten studies met the eligibility criteria, of which five studies assessed the primary outcome and six studies assessed a secondary outcome. The index tests had a 95% sensitivity and 63% specificity in diagnosing malignant from benign masses, and a 85% sensitivity and 77% specificity in diagnosing clear cell renal cell carcinoma (ccRCC) from other histological subtypes. There were no studies that assessed the diagnostic accuracy of aggressive from indolent masses. Overall, bpMRI and mpMRI demonstrated diagnostic potential, however the limited number of studies, and limitations within the study design prevent us from drawing a definitive conclusion.

Chapter 3 assessed the diagnostic value of the Liver Imaging Reporting and Data System's (LI-RADS) ancillary features (AF) in computed tomography (CT), magnetic resonance imaging (MRI) and contrast-enhanced ultrasound (CEUS). Our search strategy yielded 22 eligible studies; 5 studies assessing CT, 19 studies assessing MRI and 2 studies assessing CEUS. In our individual participant data (IPD) meta-analyses, AFs were significantly associated with malignancy, HCC or benignity, as intended by LI-RADS. These associations were apparent on multivariable analyses. In our primary analyses, 15/22 and 12/22 AFs had an independent association with HCC and malignancy, respectively. Nonenhancing capsule (OR: 3.50[95% C.I.=1.53-8.01]) and subthreshold growth (OR: 3.21[95% C.I.=1.26-8.17]) had the strongest association with increased odds of HCC, restricted diffusion (OR: 14.45[95% C.I.=9.82-21.27]) and mild-moderate T2 hyperintensity (OR: 10.18[95% C.I.=7.17-14.44]) had the strongest association with malignancy and parallels blood pool (OR: 0.02[95% C.I.=0.002-0.24) and marked T2 hyperintensity (OR: 0.06[95% C.I.=0.02-0.19]) had the strongest association with benignity. In conclusion, several AFs, in conjunction with major features (except threshold growth), could be used to diagnose liver masses in patients at high risk of HCC.

Practical Considerations and Future Research

The field of diagnostic imaging is promising, however, further research is required to determine its true impact. The conclusions drawn from chapter 2 are limited by the small sample size and high risk of bias. Furthermore, the sensitivity and specificity values are likely not high enough to replace biopsy as the diagnostic standard of care. In our review, the false positive rate for the primary outcome was 37% [95% CI:23%-54%] therefore, many benign masses were classified as malignant. In addition, for the secondary outcome ccRCC vs other subtypes, the false positive rate was 22% [95% CI:19%-27%] and false negative rate was 15% [95% CI:10%-22%]. Therefore, the histological subtype was incorrectly diagnosed in several masses. Although the evidence doesn't support using imaging alone, it may still have value if used in conjunction with biopsy and additional prognostic factors (e.g. age, comorbidities) (1). For example, the clear cell likelihood score, a 5-point scoring system that incorporates mpMRI imaging features in diagnosing ccRCC (2, 3), could be adapted to determine if biopsy is needed. Although there were only three studies included in the analysis, the clear cell likelihood score had a 96% (Range: 94-98%) sensitivity and a negative predictive value (NPV) of 92% (Range: 88-94%) if masses classified as ccLS 3-5 on MRI were considered "positive" for ccRCC (Figure 2.5, Appendix 2.10). Therefore, there is a high probability that masses classified as ccLS 1-2 are not ccRCC. Future research could utilize imaging, along with additional prognostic factors, to create a scoring system to assess the likelihood of malignancy. Patients who have a high likelihood of malignancy can be recommended for immediate surgery (no biopsy needed) and patients whose probability of malignancy remains unclear would be recommended for biopsy.

In chapter 3, several AFs had strong, independent associations to malignancy, HCC and/or benignity which emphasizes the importance of ancillary features in the diagnostic approach. The results indicate that the LI-RADS guidelines may need to be updated to better account for the impact of ancillary features. Currently, liver masses are graded by five major features (size, nonperipheral washout, enhancing capsule, threshold growth and non-rim arterial phase hyperenhancement), according to LI-RADS guidelines. The understanding of the role of AFs is limited. The application of AFs is optional, and they can only change the LI-RADS classification by one. However, in our analyses, 13/21 and 11/21 AFs had a stronger independent association with HCC and malignancy/benignity respectively, than at least one major feature (Figure 3.3, 3.4). In addition, AFs have different levels of association with HCC and malignancy, thus it may not be appropriate for them to be weighted equally (Table 3.3). Unfortunately, due to

the small sample size, there were limitations in data analysis (Please view the discussion in chapter 3 for more details) and a definitive role for AFs could not be established. Further research is needed to deduce the true association of each AF and their role in diagnosis. This can be done by adjusting for all major features and AFs, as intended by LI-RADS. Furthermore, research should look at the impact of AFs on the LI-RADS score. Ideally, all features would be compared in a single model to create an improved diagnostic method, where features would be included and individually-weighted based on their association with HCC (after adjusting for all other features). An updated model should increase the diagnostic accuracy of the LI-RADS classification system and minimize the need for invasive methods of diagnosis.

Limitations

This thesis summarized the available evidence in diagnostic imaging regarding solid renal masses and the LI-RADS' ancillary features. Although the results highlighted the potential of diagnostic imaging, there are several pitfalls that impact the internal and external validity of our reviews.

Standardization of research methodology

Firstly, the lack of non-standardized research methodology creates variability in study results. In chapter 2, the sequences and imaging parameters (Appendix 2.6) used in the index test differed between studies, increasing heterogeneity. Furthermore, the specific diagnostic methods used (Appendix 2.7) were not standardized, even between studies using the same sequences. For example, Steinberg 2021 (4) and Xu 2021a (5) used the same imaging sequences (i.e. T2-weighted, chemical shift (T1 in-phase and out-phase), dynamic contrast enhancement (DCE) and diffusion weighted imaging (DWI)) to diagnose their masses. However, different features where evaluated within each sequence to diagnose masses. Steinberg 2021 used the predefined ccLS method to diagnosis their masses, whereas Xu 2021a based their imaging results on different elements (i.e. T2 signal intensity ratio, central necrosis, apparent diffusion coefficient) in a nonspecific manner. Therefore, the comparison between studies is limited by variability in the individual study design. Future research should be aimed at developing standardized research methods for conducting and interpreting imaging results, such as the LI-RADS guidelines used in chapter 3, or the Breast Imaging and Reporting Data System (BI-RADS) used for breast

imaging (4). The benefits include the use of consistent terminology, reducing variability in imaging results and facilitating research (5).

In summary, the use of standardized research methodology will improve the comparability of primary research studies and the quality of evidence syntheses.

Data reporting and adherence to STARD guidelines

Secondly, data reporting needs to be improved in DTA studies. In chapter 2, several elements of the study design were not reported including the method of patient enrolment (7/10 studies), diagnostic criteria used (2/10 studies), the specific definition of a "solid" mass (8/10 studies) and the timing of post-contrast phases (5/10 studies) (Figure 2.2). These are potential sources of bias because it affects the value of the results and their applicability to the clinical setting. For example, if patient enrolment was not consecutive or random, the specific population may have been chosen to increase the likelihood of positive outcomes.

In chapter 3, 6/22 studies did not specify the timing between the index test and reference standard and 5/22 studies did not specify the histological diagnosis of non-HCC masses. However, due to the nature of an IPD analysis, several areas of incomplete data reporting where addressed by evaluating the raw data and directly communicating with individual study authors.

Overall, insufficient data reporting is a major issue in the medical literature because it impedes our ability to critically appraise and replicate studies (6). Furthermore, it affects our standard of care because research studies are used to guide clinical decision making by physicians, policy makers and other health workers (6). The Standards for Reporting of Diagnostic Accuracy Studies (STARD) is the current reporting guidelines for DTA studies (7). The latest version, published in 2015, was developed to *"improve the completeness and transparency of reports in diagnostic accuracy studies (7)."* It contains 30 checklist items that are necessary for readers to assess the risk of bias and applicability of DTA studies (7). However, studies have shown that adherence to the STARD guidelines is not complete. A review in 2018 by Hong et al. determined that a mean of 16.6/30 (55%) checklist items were reported across 146 DTA studies published in imaging journals in 2016. In addition, 11 items had an adherence of less than 33% (8).

In summary, reporting guidelines are used to improve the completeness and transparency of research studies (6). The research community needs to emphasize the importance of scientific rigour and thorough reporting when designing and conducting research, to improve the quality of medical literature. Furthermore, all stakeholders, including medical journals, need to be more vigilant when assessing the quality of research to ensure published articles are of high standard. This will improve the quality of downstream publications (i.e. systematic reviews, clinical practice guidelines) and decision-making in healthcare.

Quality of data in IPD analyses

Chapter 3 describes data collection and analysis of individual participant data from eligible studies. Data was evaluated using predefined methodology and variables, limiting our reliance on data reporting and outcome measures in individual studies. Critical study elements, including bias and applicability concerns, could be assessed at the observational level, using data obtained by individual authors. In addition, analyses were conducted with original data, using statistical models there were better-suited in addressing the objectives of the review. Overall, this improved the comparability between studies and increased the accuracy in our results. Although there are benefits to IPD analyses, there are significant challenges, particularly in data cleaning and validation. Initially, our IPD analysis included 28 studies, however throughout a rigorous data validation process, six studies were excluded. Two studies were excluded because the author did not want to collaborate further. The other four studies were removed because they did not assess ancillary features according to the LI-RADs guidelines.

The data from eleven of the remaining studies were modified after validation with the study authors. The main issues related to how information was coded in the data template we provided. Aside from specific features, the LI-RADs guidelines is not explicit as to what variables need to be evaluated and recorded. Therefore, when authors sent our group their data, a lot of vital information was missing or inappropriately coded. The major concern was the coding of ancillary features, as observations were either coded as “present” or “absent”. Ancillary features are optional imaging features, thus not all ancillary features are always evaluated. Furthermore, ancillary features such as size reduction of transitional phase hyperintensity are not always applicable as they depend on a prior observation or a hepatobiliary contrast agent, respectively. Observations coded as “absent” may not have been evaluated or applicable, hence

they should be excluded from the analysis. Authors had to be contacted to validate this information, amongst other concerns, to ensure that the data included in the analysis was appropriate.

To address several of the challenges IPD studies face, future research should clearly establish a data dictionary for coding data a priori, and be extremely specific. Furthermore, a multi-stage pilot extraction should be completed with a few studies to identify and address problematic areas. The first stage should be done with a homogenous group (i.e. similar expertise, imaging parameters, reference standard) to reduce inter-reader variability and determine if the study design and data extraction method are effective. The next stage should be conducted with a diverse group of authors, with different academic, professional and cultural backgrounds. A major issue in our data dictionary was that definitions/criteria that were deemed objective, were often misinterpreted by other authors. The second stage would reduce data entry errors, caused from misunderstandings, from contributing authors. In addition, researchers should review data vigilantly upon receiving it. Study authors are more active collaborators when initially sending their data and it will prevent issues from affecting future datasets. This will streamline the process of data cleaning and validation for IPD studies.

Conclusion

Biopsy is the current standard of care for diagnosing several cancers, however it is an invasive procedure and non-invasive means of diagnosis are preferable. Our thesis identified the potential value of imaging in the clinical diagnosis of solid renal masses and HCC, however the current evidence is limited. bpMRI and mpMRI could be used to minimize the need for biopsy, however further research, with standardized research methodology, is needed to reduce variability in the accuracy estimates. Several AFs were independently associated with HCC and/or malignancy, however due to limitations in the evaluation and reporting of AFs, we couldn't evaluate the true association of AFs as intended by the LI-RADS guidelines. Although there are deficits in the available evidence in diagnostic imaging, the field is new and evolving. The development of standardized research methodology and reporting guidelines, such as LI-RADS and STARD, have improved the quality of image interpretation and data reporting in DTA studies. A continued effort to create a standardized method of conducting, interpreting and

documenting information will decrease variability between study results in the future, and allow readers to determine the true impact of diagnostic imaging.

References

1. Volpe A, Patard JJ. Prognostic factors in renal cell carcinoma. World J Urol. 2010;28(3):319-27.

2. Steinberg RL, Rasmussen RG, Johnson BA, Ghandour R, De Leon AD, Xi Y, et al. Prospective performance of clear cell likelihood scores (ccLS) in renal masses evaluated with multiparametric magnetic resonance imaging. European radiology. 2021;31(1):314-24.

3. Canvasser NE, Kay FU, Xi Y, Pinho DF, Costa D, e Leon AD, et al. Diagnostic Accuracy of Multiparametric Magnetic Resonance Imaging to Identify Clear Cell Renal Cell Carcinoma in cT1a Renal Masses. The Journal of urology. 2017;198(4):780-6.

4. Steinberg RL, Rasmussen RG, Johnson BA, Ghandour R, De Leon AD, Xi Y, et al. Prospective performance of clear cell likelihood scores (ccLS) in renal masses evaluated with multiparametric magnetic resonance imaging. Eur Radiol. 2021;31(1):314-24.

5. Xu HS, Balcacer P, Zhang Z, Zhang L, Yee EU, Sun MR, et al. Characterizing T2 iso- and hypo-intense renal masses on MRI: Can templated algorithms improve accuracy? Clinical imaging. 2021;72:47-54.

6. American College of Radiology. Breast imaging reporting and data system (BI-RADS). Reston, Va.?: American College of Radiology; 1993. 64 leaves. p.

7. American College of Radiology Liver Reporting and Data System (LI-RADS) 2021 [Available from: www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/LI-RADS.

8. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527.

9. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open. 2016;6(11):e012799.

10. Hong PJ, Korevaar DA, McGrath TA, Ziai H, Frank R, Alabousi M, et al. Reporting of imaging diagnostic accuracy studies with focus on MRI subgroup: Adherence to STARD 2015. J Magn Reson Imaging. 2018;47(2):523-44.

Appendix

Appendix 2.1: Search Strategy

Ovid MEDLINE(R) ALL <1946 to January 11, 2022>

1 exp Kidney Neoplasms/ 79273
2 ((kidney or renal) adj2 (tumo?r* or carcinoma* or mass* or lesion* or neoplasm* or cancer)).tw,kf. 80453
3 rcc.tw,kf. 17347
4 or/1-3 111959
5 Magnetic Resonance Imaging/ 442622
6 exp Diffusion Magnetic Resonance Imaging/32436
7 (mri or magnetic resonance or mr imag* or diffusion weight* imag*).tw,kf. 565006
8 5 or 6 or 7 726804
9 4 and 84964
10 (sensitiv* or predictive value*).mp. or accurac*.tw. or diagnos*.ti. or predict*.tw. 4219208
11 9 and 10 1271
12 multiparametric magnetic resonance imaging/ 761
13 (((Multiparametric or multi parametric) adj3 (mri or magnetic resonance imag* or mr imag*)) or MPMRI).tw,kf. 4696
14 12 or 13 4786
15 4 and 14 44
16 11 or 15 1290
17 exp animals/ not humans/ 4941178
18 case reports.pt. 2238587
19 16 not (17 or 18) 1089

Embase Classic+Embase <1947 to 2022 January 11>

1 exp *kidney cancer/ or exp kidney carcinoma/ 102408
2 ((kidney or renal) adj2 (tumo?r* or carcinoma* or mass* or lesion* or neoplasm* or cancer)).tw. 118871
3 rcc.tw. 29276
4 1 or 2 or 3 153121
5 multiparametric magnetic resonance imaging/ 6030
6 ((multiparametric or multi parametric) adj2 (mri or magnetic resonance imag* or mr imag*)).tw. 7137
7 mpmri.tw. 3255
8 or/5-7 9077
9 4 and 8113
10 nuclear magnetic resonance imaging/ or diffusion weighted imaging/ 904536
11 (mri or magnetic resonance or mr imag* or diffusion weight* imag*).tw. 803690
12 10 or 11 1141177
13 4 and 12 9610
14 sensitiv*.tw. or diagnostic accuracy.sh,tw. or diagnostic.tw. or accura*.ti. 3086771
15 13 and 14 1881
16 9 or 15 1946

17 (exp animal/ or nonhuman/) not exp human/ 7510728
18 case report/ 2795737
19 16 not (17 or 18) 1549

EBM Reviews - Cochrane Central Register of Controlled Trials <December 2021>
1 exp Kidney Neoplasms/ 1281
2 ((kidney or renal) adj2 (tumo?r* or carcinoma* or mass* or lesion* or neoplasm* or cancer)).tw,kw. 4918
3 rcc.tw,kw. 1396
4 or/1-3 5334
5 Magnetic Resonance Imaging/ 7992
6 exp Diffusion Magnetic Resonance Imaging/433
7 (mri or magnetic resonance or mr imag* or diffusion weight* imag*).tw,kw. 41352
8 5 or 6 or 7 43026
9 4 and 8210
10 (sensitiv* or predict*).mp. or accurac*.tw. or diagnos*.ti. 216614
11 9 and 10 72
12 multiparametric magnetic resonance imaging/ 5
13 ((Multiparametric or multi parametric) adj3 (mri or magnetic resonance imag* or mr imag*)).tw,kw. 392
14 MPMRI.tw,kw. 223
15 or/12-14 463
16 4 and 15 7
17 11 or 16 73

Appendix S2.2. World Health Organization (WHO) classification of tumours of the kidney. Reproduced from the WHO International Agency for Research on Cancer and Moch et al [32]]. Morphology codes are based on the classifications from the International Classification of Diseases for Oncology (ICD-O). The morphology codes will be used to classify pathological results as benign (/0) versus malignant (/2 or /3).

Renal cell tumors		**Mesenchymal tumours occurring mainly in adults**	
Clear cell renal cell carcinoma	8310/3	Leiomyosarcoma	8890/3
Multilocular cystic renal neoplasm of low malignant potential	8316/1*	Angiosarcoma	9120/3
Papillary renal cell carcinoma	8260/3	Rhabdomyosarcoma	8900/3
Hereditary leiomyomatosis and renal cell carcinoma-associated renal cell carcinoma	8311/3*	Osteosarcoma	9180/3
Chromophobe renal cell carcinoma	8317/3	Synovial sarcoma	9040/3
Collecting duct carcinoma	8319/3	Ewing sarcoma	9364/3
Renal medullary carcinoma	8510/3*	Angiomyolipoma	8860/0
MiT family translocation renal cell carcinomas	8311/3*	Epithelioid angiomyolipoma	8860/1*
Succinate dehydrogenase-deficient renal carcinoma	8311/3	Leiomyoma	8890/0
Mucinous tubular and spindle cell carcinoma	8480/3*	Haemangioma	9120/0
Tubulocystic renal cell carcinoma	8316/3*	Lymphangioma	9170/0
Acquired cystic disease-associated renal cell carcinoma	8316/3	Haemangioblastoma	9161/1
Clear cell papillary renal cell carcinoma	8323/1	Juxtaglomerular cell tumour	8361/0
Renal cell carcinoma, unclassified	8312/3	Renomedullary interstitial cell tumour	8966/0
Papillary adenoma	8260/0	Schwannoma	9560/0
Oncocytoma	8290/0	Solitary fibrous tumour	8815/1
Metanephric tumours		**Mixed epithelial and stromal tumour family**	
Metanephric adenoma	8325/0	Cystic nephroma	8959/0
Metanephric adenofibroma	9013/0	Mixed epithelial and stromal tumour	8959/0
Metanephric stromal tumour	8935/1		
		Neuroendocrine tumours	
Nephroblastic and cystic tumours occurring mainly in children		Well-differentiated neuroendocrine tumour	8240/3
Nephrogenic rests		Large cell neuroendocrine carcinoma	8013/3
Nephroblastoma	8960/3	Small cell neuroendocrine carcinoma	8041/3
Cystic partially differentiated nephroblastoma	8959/1	Phaeochromocytoma	8700/0
Paediatric cystic nephroma	8959/0		
		Miscellaneous tumours	
Mesenchymal tumours		Renal haematopoietic neoplasms	
		Germ cell tumours	
Mesenchymal tumours occurring mainly in children			
Clear cell sarcoma	8964/3	**Metastatic tumours**	
Rhabdoid tumour	8963/3		
Congenital mesoblastic nephroma	8960/1		
Ossifying renal tumour of infancy	8967/0		

Appendix 2.3. Histologic classification of aggressive versus indolent renal tumors [reproduced from Bhindi et al [17]]. This table will be used to classify pathological results as indolent versus aggressive.

Indolent[a]	**Aggressive** a
Benign and indolent	Aggressive clear-cell RCC c
Oncocytoma	Aggressive papillary RCC c
Non-epithelioid angiomyolipoma [b]	Collecting duct RCC
Papillary adenoma	Unclassified RCC
Metanephric adenoma	Translocation-associated RCC
Other benign tumors	Hereditary leiomyomatosis RCC
	Other malignant non-RCC tumors
Malignant and indolent	
Epithelioid angiomyolipoma b	
Indolent clear-cell RCC c	
Indolent papillary RCC c	
Chromophobe RCC [d]	
Clear-cell papillary RCC	
Tubulocystic RCC	
Mucinous tubular and spindle cell RCC	
Succinyl dehydrogenase deficient RCC	

RCC = renal cell carcinoma.

a. The presence of any sarcomatoid differentiation led to the classification of malignancies as aggressive. The presence of coagulative necrosis also led to the classification of malignancies as aggressive, except for low-grade (1–2) papillary RCC, where it does not appear to portend a poor prognosis [26].

b. Epithelioid angiomyolipoma was considered malignant given its ability to metastasize, but indolent in behavior [33], [34].

c. In addition to sarcomatoid differentiation (both clear-cell and papillary RCC) and necrosis (clear-cell RCC), any high-grade (3–4) component led to aggressive classification for clear-cell and papillary RCC.

d. Chromophobe RCC was not graded as per International Society of Urological Pathology 2012 Consensus Recommendations [29] and was considered indolent if sarcomatoid differentiation and coagulative necrosis were absent.

Appendix 2.4. Modified QUADAS-2 signalling questions for risk of bias and applicability assessment.

Domain 1: Patient selection A. Risk of bias Describe methods of patient selection: I. Was a consecutive or random sample of patients enrolled? II. Was a case-control design avoided? III. Did the study avoid inappropriate exclusions Could the selection of patients have introduced bias? B. Concerns regarding applicability Is there concern that the included patients do not match the review question? *I. If the study inclusion includes all renal masses but does not specify solid vs. cystic proportion = HIGH (there is a high likelihood that cystic renal masses were included)* *II. If the study definition of solid masses includes masses with >25% cystic component = HIGH* III. *If the study includes solid masses but does not specify the definition of "solid" = UNCLEAR*	 Yes, no or unclear Yes, no or unclear Yes, no or unclear Risk: low, high or unclear Concern: low, high or unclear
Domain 2: Index test and comparator test A. Risk of bias Describe the index test and how it was conducted and interpreted: a. Were the index test results interpreted without knowledge of the results of the reference standard? b. Were diagnostic criteria prespecified, including any thresholds used for quantitative parameters? *i. MODIFIED from "If a threshold was used, was it prespecified"* Could the conduct or interpretation of the index test have introduced bias? B. Concerns regarding applicability Is there concern that the index test, its conduct, or interpretation differ from the review question? *I. If the DCE timing does not adhere to the recommendations from the Society of Abdominal Radiology Disease Focused Panel on Renal Cell Carcinoma (23) (i.e: corticomedullary phase (~30 seconds post contrast), nephrographic phase (~90-100 seconds post contrast) ± excretory phase (180-210 seconds post contrast) imaging = HIGH* II. *If timing of post-contrast sequences not sufficiently reported = UNCLEAR*	 Yes, no or unclear Yes, no or unclear Risk: low, high or unclear Concern: low, high or unclear
Domain 3: Reference standard A. Risk of bias Describe the reference standard and how it was conducted and interpreted:	

I. Is the reference standard likely to correctly classify the target condition? *i. Any histopathology (surgical or biopsy) = YES* Could the reference standard, its conduct, or its interpretation have introduced bias? B. Concerns regarding applicability Is there concern that the target condition as defined by the reference standard does not match the review question? *I. Any histopathology (surgical or biopsy) = LOW*	Yes, no or unclear Risk: low, high or unclear Concern: low, high or unclear
<u>Domain 4: Flow and timing</u> A. Risk of bias Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table: Describe the time interval and any interventions between index test(s) and reference standard: I. Did all patients receive a reference standard? II. Did patients receive the same reference standard? III. Were all patients included in the analysis? Could the patient flow have introduced bias?	 Yes, no or unclear Yes, no or unclear Yes, no or unclear Risk: low, high or unclear

Appendix 2.5. Specific subtypes of solid renal masses used in each study.

Study	Outcome	Number of masses	Target condition	Comparator group
Xu 2021a	Malignant vs. benign	60	ccRCC, pRCC; chrRCC, "malignant others" (RCC, RCC with oncocytic and papillary features, other renal cell carcinoma subtypes including Xp11 translocation RCC, mixed histopathology tumors including hybrid oncocytoma and chromophobe tumor (HOCT))	AML, Oncocytoma, a nonspecified mesenchymal tumor that could not be further differentiated by the pathologist
Goyal 2018	Malignant vs. benign	90	RCC, TCC, Miscellaneous (SCC, Primitive neuroectodermal tumour, Lymphoma, metastases)	AML, oncocytoma, leiomyoma
Baldari 2015	Malignant vs. benign; ccRCC vs other subtypes	18	ccRCC, ccpRCC, urothelial carcinoma, chrRCC, lymphoma*	lipoma, oncocycoma
Xu 2021b	Malignant vs. benign	44	ccRCC, pRCC, chrRCC	fp-AML, oncocytoma
De Silva 2021	Malignant vs. benign	72	Radiologist impression	Radiologist impression
Steinberg 2021	ccRCC vs other subtypes	454	pRCC, chrRCC, oncocytoma, oncocytic neoplasm; unspecified RCC, urothelial cell carcinoma, primary renal sarcoma, renal medullary carcinoma, unknown carcinoma, metastatic lesion, benign	mpMRI
Nikpanah 2021	ccRCC vs other subtypes	80 included in radiologist subset (of 243 total)	ccRCC	Oncocytoma
Cupido 2017	ccRCC vs other subtypes	50	ccRCC	pRCC, chrRCC
Canvasser 2017	ccRCC vs other subtypes	121	ccRCC	All other subtypes (pRCC, chrRCC, other RCC, oncocytoma, AML, other benign)
Schieda 2022	ccRCC vs other subtypes	250	ccRCC	pRCC, oncocytoma, chrRCC, fpAML, unclassified RCC, ccpRCC, oncocytic neoplasm, metastasis, mucinous tubular and spindle cell tumor, epithelioid AML, lymphoma, multilocular cystic neoplasm of low malignant potential, sclerosing PEComa

RCC= renal cell carcinoma; ccRCC=clear cell renal cell carcinoma; pRCC=papillary renal cell carcinoma; chrRCC=chromophobe renal cell carcinoma; AML=angiomyolipoma; TCC=renal transitional cell carcinoma; SCC=squamous cell carcinoma; ccpRCC=clear cell papillary renal cell carcinoma; fp-AML=fat poor angiomyolipoma;
*For the study by Baldari et al.[48], ccpRCC, urothelial carcinoma, chrRCC and lymphoma masses were part of the comparator group when assessing the secondary outcome ccRCC vs other subtypes.

Appendix 2.6. Detailed description of MRI parameters used in imaging of renal masses (NR=not reported)

Study	Instrument	Vendor	Sequence	Specific Sequence	Repetition time (ms)	Echo time (ms)	Flip angle (degrees)	Slice thickness (mm)	Matrix size	Field of view (cm)
Xu 2021a	1.5T/3.0T MRI Scanner	Symphony or Espree Siemens Medical Solutions or Signa Excite TwinSpeed, GE Medical Systems/ Signa Excite, GE Medical Systems	*T2 weighted*	T2 Coronal fast spin echo	80-1000	60-70	90	4-5	256x192 (1.5T) 288x192 (3.0T)	38-42
			Chemical shift (T1 in-phase and out-phase)	T1 Axial dual-echo in and opposed phases, spoiled gradient recall echo (GRE)	170-80	2.2-2.8/4.4-5.3 (1.5T) 1.1-1.3/2.2-2.6 (3.0T)	80	6-7	256x128-192 (1.5T) 320x160 (3.0T)	36-40
			Dynamic contrast enhancement	Coronal/Axial 3D GRE with fat saturation, post contrast	4-5	1-2	10-11/11-15	3/2.5-3.2	256x192 (1.5T) 256x160 (3.0T)/ 256x192	38-42/36-40
			Other (please specify)	Sagittal oblique (oriented across each kidney) 3D GRE with fat saturation	4-5	1-2	10 (1.5T) 12 (3.0T)	3-3.2	256x192 (1.5T) 256x160 (3.0T)	34-38
Steinberg 2021	1.5T/3.0T whole-body scanner	Philips Ingenia 1.5T and 3T, Philips Achieva 1.5T and 3T, Siemens Avanto 1.5T, and Siemens Aera 1.5T; institution 2: Siemens Aera 1.5T, Siemens Avanto Fit 1.5T, Siemens Skyra 3T)	*T2 weighted*	NR	NR	NR	NR	NR	NR	NR
			Chemical shift (T1 in-phase and out-phase)	NR	NR	NR	NR	NR	NR	NR
			Dynamic contrast enhancement	NR	NR	NR	NR	NR	NR	NR
			DWI (optional)	NR	NR	NR	NR	NR	NR	NR
Nikpanah 2021	1.5T MRI Scanner	Aera, Siemens Healthcare	*T2 weighted*	T2-WI without fat saturation	1700	112	140	40	232x256	148.5x66.9
			Chemical shift	Pre-contrast T1-WI	3.78	1.74	10	104	270x320	119x53.5

			Dynamic contrast enhancement	Arterial Phase T1-WI/ Venous Phase T1-Wi	3.78	1.74	10	104	270x320	119x53.5
Goyal 2018	1.5T MRI Scanner	Siemens, Avanto, Erlangen	*T2 weighted*	True FISP 2D (Axial/Coronal) T2W FS TSE (Axial/Coronal)	3.4/3.4 2520/2700	1.4/1.4 100/100	NR	5/5 5/5	288x512/410x512 288x512/171x256	263x350/380x380 278x370/410x430
			Chemical shift (T1 in-phase and out-phase)	T1W 2D FLASH axial dual echo (In-phase/Out-of-phase)	125/125	4.76/2.34	NR	5/5	288x512/288x512	278x370/278x370
			Dynamic contrast enhancement	Post contrast T1W FS 3D VIBE (Axial/Coronal)	5.1/5.0	2.3/2.3	NR	3/3	158x512/322x512	253x450/400x400
			DWI (optional)	FS axial DW-MRI at b=0,500 s/mm^2	1600	62	NR	7	94x192	249x380
Cupido 2017	1.5T MRI Scanner	Siemens Healthcare, Erlangen, Germany	*T2 weighted*	Fast imaging with steady-state free precession (FISP) sequence acquired coronally; Half-Fourier acquisition single-shot turbo spin-echo (HASTE) sequence acquired axially; Fat-suppressed HASTE sequence acquired coronally	3.6; 1000; 900	1.83; 90; 90	70; NR; NR	6; 6; 6	NR	NR
			Chemical shift (T1 in-phase and out-phase)	Dual echo T1 in-phase and T1-opposed phase sequences acquired axially	136/136	4.7/2.03	70/70	6/6	NR	NR
			Dynamic contrast enhancement	Fat-suppressed unenhanced and contrast-enhanced three-dimensional (3D) volumetric interpolated breath-hold examination (VIBE) sequences acquired axially; Fat-suppressed contrast-enhanced 3D VIBE sequence acquired axially	4.3; 4.3	1.68; 1.68	120; 120	4; 4	90x9x256; NR	225x9x450; NR
			DWI (optional)	Diffusion-weighted navigator-triggered prospective acquisition correction (DWI-PACE) with b values of 0, 50, 150, and 500, acquired axially.	NR	NR	NR	NR	NR	NR
Canvasser 2017	1.5T/3.0T whole-body scanner	NR	*T2 weighted*	Coronal/Axial (with fat-suppression) planes	964-4820	70-120	NR	4-8	256x154 to 348x280	NR
			Chemical shift (T1 in-phase and out-phase)	Axial chemical shift T1-weighted images	100-216	2.38/4.87 (1.5T) 1.12.3 (3.0T)	NR	5-8	193x168 to 500x286	NR
			Dynamic contrast	Fat-supressed DCS T1-weighted imaging (Axial or coronal planes)	3.4-5.4	1.6-2.4	NR	3-6	176x149 to 320x259	NR

			enhancement							
			DWI (optional)	Diffusion-weighted navigator-triggered prospective acquisition correction (DWI-PACE) with b values of 0, 50, 150, and 500, acquired axially	NR	NR	NR	NR	NR	NR
Baldari 2015	High-field (3 Tesla) scanner	Trio, Siemens	*T2 weighted*	T2 weighted HASTE axial and coronal planes with and without fat suppression/ T2 weighted TRUFI on axial and coronal planes	2000/ 510.3	92/ 1.2	NR	NR	NR	NR
			Chemical shift (T1 in-phase and out-phase)	Axial chemical shift T1-weighted images/ T1 weighted VIBE on the axial plane	1500 (In-phase); 1530 (Out-of-phase)/ 3.3	2 (In-phase); 35 (Out-of-phase)/ 1.1	NR	NR	NR	NR
			Dynamic contrast enhancement	Axial plane DCE	3.3	1.1	NR	NR	NR	NR
Xu 2021b	High-field (3 Tesla) scanner	Trio, Siemens	*T2 weighted*	Axial T2-weighted fast spin-echo sequence (T2WI)	4800	74	NR	5/1	320x320	40x40
			DWI (optional)	Breath-hold single-shot spin echo echo-planar imaging (SE-EPI) (b-value=800 seconds/mm)	7000	80	NR	5/1.5	NR	24x24
De Silva 2021	High-field (3 Tesla) scanner	NR	*T2 weighted*	Conventional breath-held sequences	NR	NR	NR	NR	NR	NR
			Chemical shift (T1 in-phase and out-phase)	NR	NR	NR	NR	NR	NR	NR
			Dynamic contrast enhancement	NR	NR	NR	NR	NR	NR	NR
			DWI (optional)	NR	NR	NR	NR	NR	NR	NR
Schieda 2022	1.5T/3.0T whole-body scanner	1.5T: Symphony or Aera, Siemens Healthcare 3.0T: Trio, Siemens Healthcare; Discovery 750 W, General	*T2 weighted*	Axial and/or coronal T2-weighted (T2W) single-shot fast or turbo spin echo	All imaging sequences were performed and met study inclusion criteria. A full detail of the individual parameters can be accessed at: (INSERT LINK)					
			Chemical shift (T1 in-phase and out-phase)	Dual-echo chemical shift (in and opposed phase) T1-weighted (T1W) gradient-recalled echo (GRE)						
			Dynamic contrast enhancement	Multiphasic contrast-enhanced fat-suppressed (FS) T1W GRE obtained before and after intravenous administration						

		Electric Healthcare	*DWI (optional)*	Axial single-shot echo-planar diffusion-weighted imaging (DWI) with a low (≤200 mm2/sec) and high (≥500 mm2/sec) b-value	

NR-Not reported; DWI- Diffusion weighted imaging

Appendix 2.7. MRI features used for diagnosis of target condition among renal masses

Study	T2 weighted		Chemical shift (T1 in-phase and out-phase)		Dynamic contrast enhancement		DWI (optional)		Other (please specify)		Primary diagnostic criteria
	Qualitative	Quantitative	Qualitative	Quantitative	Qualitative	Quantitative	Qualitative	Quantitative	Qualitative	Quantitative	
Xu 2021a	Heterogeneity of the lesion -- Lesion texture as observed on T2WI; Central necrosis (T2 bright nonenhancing))	T2 Signal Intensity Ratio (T2 SIR) -- Ratio of tumor to normal renal cortex SI on non-fat-saturated T2 weighted images (coronal or axial)	Intravoxel fat (signal drop on OOP); Hemorrhagic component (T1 hyperintense without enhancement); Central necrosis (T1 dark nonenhancing)		Segmental Enhancement Inversion (SEI) or T2 bright delayed enhancing central scar.	Arterial-to-Delay enhancement Ratio (ADER), cutoff 1.5;		ADC (cutoff of 1.3 and 1.7); macroscopic fat without calcification	Macroscopic fat without calcification (signal dropout on fat suppressed images)		Undefined

Steinberg 2021	Signal intensity on T2 vs Renal Cortex (ccRCC=hyperintense/isotense/hypointense; pRCC=isotense/hypointense; chrRCC=hyperintense/isotense; AML=hypointense; oncocytoma=hyperintense/isotense)		Microscopic fat (present or not present); Central non-enhancing area and/or Intravoxel fat (ccRCC=present)		Determine enhancing component on delay (2min) post-contrast; Assess area with highest enhancement on CM phase (ccRCC=intense; pRCC=mild; chrRCC=intense/moderate; AML=intense/mild; oncocytoma=intense/moderate); segmental enhancement inversion (present or not present); Central scar (ccRCC and oncocytoma=present)	Arterial-to-Delay enhancement Ratio (ADER), cutoff 1.5 (ccRCC <1.5; AML>1.5)	Diffusion restriction (higher and lower signal intensity than renal parenchyma on b800 DWI and ADC, respectively)		Macroscopic fat (ccLS is not applicable to masses with macroscopic fat)		Clear cell likelihood score (ccLS)
Nikpanah 2021											Undefined

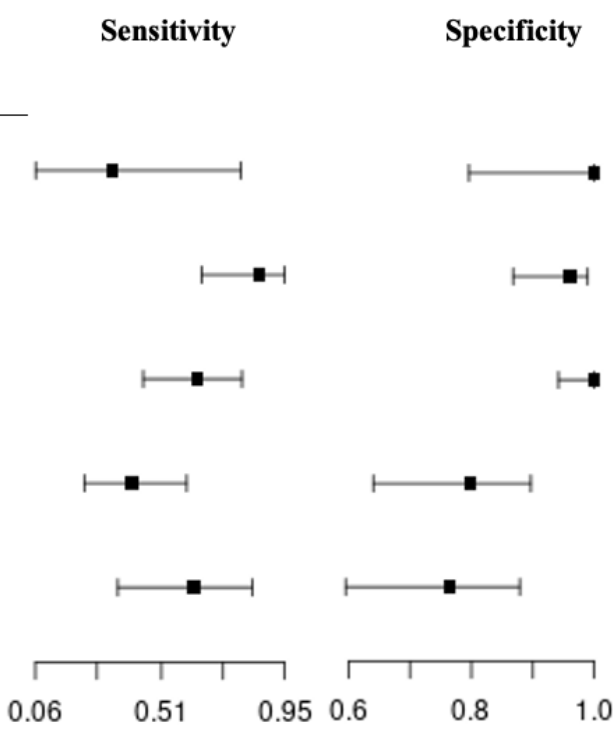
Goyal 2018							FS axial DW-MRI at b=0, 500 s/mm2; presence of restricted diffusion in the solid enhancing components of a renal lesion affirmed malignancy; Free diffusion in cystic lesions clinched the diagnosis of benign cysts				Undefined

Cupido 2017	T2 signal on MRI (ccRCC=mostly high signal; pRCC=mostly low signal; chrRCC=variable signal)				Tumor hypervascularity on contrast enhanced images (ccRCC=hypervascular; pRCC=hypovascular; chrRCC=intermediate vascularity)				Margins (ccRCC=well or poorly circumscribed; pRCC=well circumscribed; chrRCC=well circumscribed and relatively large) Consistency (ccRC=mostly heterogenous; pRCC=mostly homogenous; chrRCC-central stellate scar, but if no scar, mostly homogenous)		Margins, consistency, T2 signal and vascularity

Canvasser 2017	Signal intensity on T2 vs Renal Cortex (ccRCC=hyperintense/isotense/hypointense; pRCC=isotense/hypointense; chrRCC=hyperintense/isotense; AML=hypointense; oncocytoma=hyperintense/isotense)		Microscopic fat (present or not present); Central non-enhancing area and/or Intravoxel fat (ccRCC=present)		Determine enhancing component on delay (2min) post-contrast; Assess area with highest enhancement on CM phase (ccRCC=intense; pRCC=mild; chrRCC=intense/moderate; AML=intense/mild; oncocytoma=intense/moderate); segmental enhancement inversion (present or not present); Central scar (ccRCC and oncocytoma=present)	Arterial-to-Delay enhancement Ratio (ADER), cutoff 1.5 (ccRCC <1.5; AML>1.5)	Diffusion restriction (higher and lower signal intensity than renal parenchyma on b800 DWI and ADC, respectively)		Macroscopic fat (ccLS is not applicable to masses with macroscopic fat)		Undefined
Baldari 2015	Presence or absence of changes of the renal parenchyma in terms of space-occupying lesions	Mass size			Lesion contrast enhancement				Margins; Peri-renal fat infiltration; Loco-regional vascular structures infiltration; lesion structure; overall not very specific		Undefined
Xu 2021b	Presumed. no details provided						Presumed. no details provided				Undefined

De Silva 2021			Microscopic fat (ccRCC=common; pRCC=uncommon; chrRCC=common; lipid poor AML=uncommon/rare; fat rich AML=common; oncocytoma=rare/never), macroscopic fat (fat rich AML=always; others=none)		Enhancement ratio in corticomedullary phase(ccRCC=avid; pRCC=mild; chrRCC=moderate-avid; lipid poor AML=mild-moderate; fat rich AML=variable; oncocytoma=avid)			ADC values (ccRCC=intermediate 1.43, (0.68-2.46); pRCC=low(0.77, 0.55-0.90); chrRCC=low(0.66, 0.66-0.66); lipid poor AML=low (1.09, 1.05-1.20) ; fat rich AML=low (0.74, 0.48-0.83) ; oncocytoma=high (2.16, 1.47-2.59))	Necrosis (ccRCC=common; chrRCC=rare; fat rich AML=rare; oncocytoma=rare/may have scar)		ADC, microscopic and macroscopic fat, enhancement ratio in CM phase, necrosis (if applicable)
Schieda 2022	Signal intensity on T2 vs Renal Cortex (ccRCC=hyperintense/isotense/hypointense; pRCC=isotense/hypointense; chrRCC=hyperintense/isotense; AML=hypointense; oncocytoma=hyperintense/isotense)		Microscopic fat (present or not present); Central non-enhancing area and/or Intravoxel fat (ccRCC=present)		Determine enhancing component on delay (2min) post-contrast; Assess area with highest enhancement on CM phase (ccRCC=intense; pRCC=mild; chrRCC=intense/moderate; AML=intense/mild; oncocytoma=intense/moderate); segmental enhancement inversion (present or not present); Central scar (ccRCC and oncocytoma=present)	Arterial-to-Delay enhancement Ratio (ADER), cutoff 1.5 (ccRCC <1.5; AML>1.5)	Diffusion restriction (higher and lower signal intensity than renal parenchyma on b800 DWI and ADC, respectively)		Macroscopic fat (ccLS is not applicable to masses with macroscopic fat)		Clear cell likelihood score (ccLS)

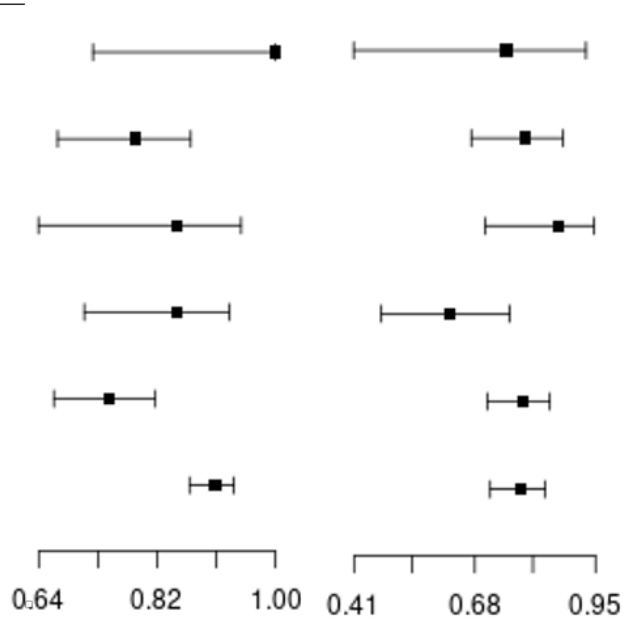
Study	TP	FP	FN	TN	Sensitivity [95 C.I.]	Specificity [95 C.I.]
Baldari 2015	15	2	0	1	1.00 [0.80, 1.00]	0.33 [0.06, 0.79]
De Silva 2021	49	3	2	18	0.96 [0.87, 0.99]	0.86 [0.65, 0.95]
Goyal 2018	62	9	0	16	1.00 [0.94, 1.00]	0.64 [0.45, 0.80]
Xu 2021a*[a]	29	14	7	10	0.80 [0.64, 0.90]	0.40 [0.23, 0.60]
Xu 2021b*[b]	25	4	7	8	0.77 [0.60, 0.88]	0.62 [0.35, 0.84]
Summary Estimate					***0.95 [0.77, 0.99]***	***0.63 [0.46, 0.77]***



*Data was averaged out over several readers.
[a]Fleiss's kappa (three readers): r=0.29 (p=0.0001)
[b]Inter-class correlation coefficient between radiologist A and B: 0.822

Appendix 2.8. Individual two-by-two data and forest plots of the included studies for the primary target condition: malignant vs. benign solid renal masses (n=5). Summary estimates of sensitivity and specificity, with 95% CIs, were calculated using the bivariate random effects model. The figure was generated using the lme4 package in R.

Study	TP	FP	FN	TN	Sensitivity [95 C.I.]	Specificity [95 C.I.]
Baldari 2015	10	2	0	6	1.00 [0.72, 1.00]	0.75 [0.41, 0.93]
Canvasser 2017*[ab]	48	13	12	46	0.79 [0.67, 0.87]	0.79 [0.67, 0.88]
Cupido 2017*	17	3	4	26	0.85 [0.64, 0.95]	0.87 [0.70, 0.95]
Nikpanah 2021	34	15	6	25	0.85 [0.71, 0.93]	0.62 [0.47, 0.76]
Schieda 2022*[bc]	89	30	28	103	0.75 [0.66, 0.82]	0.79 [0.71, 0.85]
Steinberg 2021[b]	258	37	26	133	0.91 [0.87, 0.94]	0.78 [0.71, 0.84]
Summary Estimate					***0.85 [0.77, 0.90]***	***0.77 [0.73, 0.81]***



*Data was averaged out over several readers.
[a]Inter-reader variability (seven readers): κ=0.53 (range: 0.38-0.64)
[b]Studies used ccLS for diagnosis. Masses were defined as ccRCC "positive" if they had score of 4 or 5 [18, 43], while other masses (ccLS 1-3) were classified as "negative."
[c]Fleiss's kappa- Institution 1 (2 readers) r=0.46 (95% CI=0.28-0.63); Institution 2 (2 readers) r=0.58 (95% CI=0.39-0.77); Institution 3 (2 readers) r=0.66 (95% CI=0.46-0.85); Institution 4 (2 readers) r=0.41 (95% CI=0.26-0.60); Institution 5 (2 readers) r=0.82 (95% CI=0.64-1.00);

Appendix 2.9. Individual two-by-two data and forest plots of the included studies for the secondary target condition: ccRCC vs other subtypes (n=6). Summary estimates of sensitivity and specificity, with 95% CIs, were determined using the bivariate random effects model. The figure was generated using the lme4 package in R.

Appendix 2.10. Individual study data for the diagnoses of ccRCC using ccLS scores (n=3). Steinberg 2021 [18], Canavsser 2017 [43] and Schieda 2022 [47] used ccLS scores to diagnose ccRCC. The ccLS scores presented serve as thresholds (ex. ccLS 3 classifies a ccRCC positive mass as any mass diagnosed as ccLS 5, ccLS 4 or ccLS 3).

ccLS Score	Study	TP	FP	FN	TN	Sensitivity [95 C.I.]	Specificity [95 C.I.]
5	Steinberg 2021	173	13	111	157	0.61 [0.55, 0.66]	0.92 [0.87, 0.95]
	Canvasser 2017*[a]	23.6	4.3	37	54.1	0.39 [0.28, 0.52]	0.93 [0.83, 0.97]
	Schieda 2022*[b]	45	10.5	74	120.5	0.38 [0.30, 0.47]	0.92 [0.86, 0.96]
4	Steinberg 2021	258	37	26	133	0.91 [0.87, 0.94]	0.78 [0.71, 0.84]
	Canvasser 2017*[a]	47.3	12.2	13.3	46.2	0.78 [0.66, 0.87]	0.79 [0.67, 0.88]
	Schieda 2022*[b]	89	28.5	30	102.5	0.75 [0.66, 0.82]	0.78 [0.70, 0.84]
3	Steinberg 2021	278	74	6	96	0.98 [0.95, 0.99]	0.56 [0.49, 0.64]
	Canvasser 2017*[a]	57.9	25.3	2.7	33.1	0.96 [0.87, 0.99]	0.57 [0.44, 0.69]
	Schieda 2022*[b]	112	78	7	53	0.94 [0.88, 0.97]	0.40 [0.32, 0.49]
2	Steinberg 2021	281	118	3	52	0.99 [0.97, 1.00]	0.31 [0.24, 0.38]
	Canvasser 2017*[a]	60.2	36.3	0.4	22.1	0.99 [0.93, 1.00]	0.38 [0.27, 0.51]
	Schieda 2022*[b]	116	84.5	3	46.5	0.97 [0.93, 0.99]	0.35 [0.28, 0.44]
1	Steinberg 2021	284	170	0	0	1.00 [0.98, 1.00]	0.00 [0.00, 0.03]
	Canvasser 2017*[a]	60.6	58.4	0	0	0.99 [0.93, 1.00]	0.01 [0.00, 0.08]
	Schieda 2022*[b]	119	131	0	0	1.00 [0.96, 1.00]	0.00 [0.00, 0.04]

*Data was averaged out over several readers.

[a]Inter-reader variability (seven readers): κ=0.53 (range: 0.38-0.64)

[b]Fleiss's kappa- Institution 1 (2 readers) r=0.46 (95% CI=0.28-0.63); Institution 2 (2 readers) r=0.58 (95% CI=0.39-0.77); Institution 3 (2 readers) r=0.66 (95% CI=0.46-0.85); Institution 4 (2 readers) r=0.41 (95% CI=0.26-0.60); Institution 5 (2 readers) r=0.82 (95% CI=0.64-1.00);

Appendix 3.1. Composite Reference Standard (12)

For HCC, histopathology from core needle biopsy, hepatectomy, or explanation was favored. Otherwise, an observation was considered positive for HCC if it fulfilled either of the following criteria:

- Categorized as LR-5 on another imaging modality and showed threshold growth (≥ 50% size increase in less than 6 months);
- Categorized as LR-5 and treated locoregionally, then recurred on CT or MRI based on LI-RADS treatment response criteria.

Non-HCC malignancies required histopathology for confirmation. LR-3, LR-4 and LR-M observations that underwent local treatment and recurred on subsequent CT or MRI were considered malignant, but not specifically HCC.

An observation was considered benign if it was stable on imaging for at least 12 months, or if it demonstrated either an unequivocal spontaneous size reduction of at least 30% diameter or disappearance with no treatment, and not attributable to resorption of tumoral blood products.

Appendix 3.2. Search strategy used in this review. The search strategy was adapted from the previous IPD study (13).

Initial Search

Database: Ovid MEDLINE(R) ALL <1946 to January 12, 2022>
Search Strategy:

Search Statement Results
1 (LI-RADS or LIRADS).tw,kw,kf. 213
2 "liver imaging reporting and data system."tw,kw. 164
3 (LR-1 or LR-2 or LR-3 or LR-4 or LR-5 or LR-5 V or LR-OM or LR-TIV or LR-M or LR1 or LR2 or LR3 or LR4 or LR5 or LR5 V or LRTIV or LRM or LROM).tw.
1381
4 or/1–3 1535
5 Radiology Information Systems/ 5601
6 Carcinoma, Hepatocellular/or Cholangiocarcinoma/ 87260
7 (hepatocellular carcinoma* or hepatocellular neoplasm* or hepatocellular cancer or hepatic cell carcinoma* or HCC or cholangiocarcinoma* or hepatic nodule* or liver lesion* or adrenocortical carcinoma*).tw,kw,kf.
106476
8 Liver Neoplasms/dg 14552
9 exp Liver Diseases/dg or liver disease*.tw,kw,kf. 125220
10 Liver/dg [Diagnostic Imaging] 14373
11 liver imaging.tw,kw. 948
12 or/6–11 244663
13 5 and 12 63
14 4 or 13 1570
15 Tomography, X-Ray Computed/ 366371
16 (computed tomograph* or CT).tw,kw. 465130
17 Magnetic Resonance Imaging/ 382106
18 (mri or magnetic resonance or MDCT).tw,kw,kf. 459570
19 Ultrasonography, Doppler, Color/ 13621
20 (CEUS or contrast enhanced ultrasound).tw,kw. 4247
21 or/15–20 1131135
22 14 and 21 286
23 limit 22 to yr = "2014-Current" 234

Database: Embase Classic+Embase <1947 to 2022 January 12>
Search Strategy:

1 (LI-RADS or LIRADS).tw. (296)
2 "Liver Imaging Reporting and Data System".af. (230)
3 (LR-1 or LR-2 or LR-3 or LR-4 or LR-5 or LR-5V or LR-OM or LR-TIV or LR-M or LR1 or LR2 or LR3 or LR4 or LR5 or LR5V or LRTIV or LRM or LROM).tw. (2063)
4 or/1-3 (2277)
5 computer assisted tomography/(703333)

6 (computed tomograph* or ct).tw. (714930)
7 nuclear magnetic resonance imaging/(745564)
8 (MRI or magnetic resonance or mr imaging or MDCT).tw. (679741)
Page 6 of 34
9 contrast-enhanced ultrasound/(3100)
10 (CEUS or contrast enhanced ultrasound).tw. (6818)
11 or/5-10 (1788445)
12 4 and 11 (426)
13 limit 12 to yr="2014 -Current" (352)

Database: EBM Reviews - Cochrane Central Register of Controlled Trials <December 2021>
Search Strategy:
1 (LI-RADS or LIRADS).tw,kw (5).
2 "liver imaging reporting and data system".tw,kw (4).
3 (LR-1 or LR-2 or LR-3 or LR-4 or LR-5 or LR-5V or LR-OM or LR-TIV or LR-M or LR1 or LR2 or LR3 or LR4 or LR5 or LR5V or LRTIV or LRM or LROM).tw. (392)
4 or/1-3 (396)
5 Radiology Information Systems/ (25)
6 Carcinoma, Hepatocellular/or Cholangiocarcinoma/(1718)
7 (hepatocellular carcinoma* or hepatocellular neoplasm* or hepatocellular cancer or hepatic cell carcinoma* or HCC or cholangiocarcinoma* or hepatic nodule* or liver lesion* or adrenocortical carcinoma*).tw,kw. (5710)
8 Liver Neoplasms/(2229)
9 exp Liver Diseases/or liver disease*.tw,kw. (19094)
10 Liver/(2960)
11 liver imaging.tw,kw (51).
12 or/6-11 (24377)
13 5 and 12 (1)
14 4 or 13 (397)
15 Tomography, X-Ray Computed/(4051)
16 (computed tomograph* or CT).tw. (29214)
17 Magnetic Resonance Imaging/(6879)
18 (mri or magnetic resonance or MDCT).tw. (28896)
19 Ultrasonography, Doppler, Color/(440)
20 (CEUS or contrast enhanced ultrasound).tw. (280)
21 or/15-20 (56293)
22 14 and 21 (31)
Page 7 of 34
23 limit 22 to yr="2014 -Current" (24)

SCOPUS
Search Strategy:
((TITLE-ABS-KEY (li-rads OR lirads)) OR (TITLE-ABS-KEY ("liver imaging reporting and data system")) OR (TITLE-ABS-KEY (lr-1 OR lr-2 OR lr-3 OR lr-4 OR lr-5 OR lr-5v OR lr-om OR lr-tiv OR lr-m OR lr1 OR lr2 OR lr3 OR lr4 OR lr5 OR lr5v OR

lrtiv OR lrm OR lrom))) AND ((TITLE-ABS-KEY ("computed tomograph*" OR ct)) OR (TITLE-ABS-KEY (mri OR magnetic AND resonance OR mdct)) OR (TITLE-ABS-KEY (ceus OR contrast AND enhanced AND ultrasound))) AND (LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019))

Updated Search

Embase Classic+Embase <1947 to 2022 January 12>
Ovid MEDLINE(R) ALL <1946 to January 12, 2022>
EBM Reviews - Cochrane Central Register of Controlled Trials <December 2021>
Search Strategy:

1 (LI-RADS or LIRADS).tw,kw,kf. 1203
2 "liver imaging reporting and data system".tw,kw. 820
3 (LR-1 or LR-2 or LR-3 or LR-4 or LR-5 or LR-5V or LR-OM or LR-TIV or LR-M or LR1 or LR2 or LR3 or LR4 or LR5 or LR5V or LRTIV or LRM or LROM).tw. 5129
4 1 or 2 or 3 5901
5 Radiology Information Systems/ 6755
6 Carcinoma, Hepatocellular/ or Cholangiocarcinoma/208970
7 (hepatocellular carcinoma* or hepatocellular neoplasm* or hepatocellular cancer or hepatic cell carcinoma* or HCC or cholangiocarcinoma* or hepatic nodule* or liver lesion* or adrenocortical carcinoma*).tw,kw,kf. 337880
8 Liver Neoplasms/dg 17086
9 exp Liver Diseases/dg or liver disease*.tw,kw,kf. 149105
10 Liver/dg 17189
11 liver imaging.tw,kw. 3255
12 or/6-11527939
13 5 and 12 119
14 4 or 13 5943
15 Tomography, X-Ray Computed/ 482737
16 (computed tomograph* or CT).tw,kw. 1470658
17 Magnetic Resonance Imaging/ 1034374
18 (mri or magnetic resonance or MDCT).tw,kw,kf. 1410630
19 Ultrasonography, Doppler, Color/ 23727
20 (CEUS or contrast enhanced ultrasound).tw,kw. 14883
21 or/15-20 3186872
22 14 and 21 1451
23 22 use medall 594
24 limit 23 to dt=20190912-20220113 302
25 (LI-RADS or LIRADS).tw. 1154
26 "Liver Imaging Reporting and Data System".af. 1013
27 (LR-1 or LR-2 or LR-3 or LR-4 or LR-5 or LR-5V or LR-OM or LR-TIV or LR-M or LR1 or LR2 or LR3 or LR4 or LR5 or LR5V or LRTIV or LRM or LROM).tw. 5129
28 or/25-27 5931
29 computer assisted tomography/ 803640
30 (computed tomograph* or ct).tw. 1454231
31 nuclear magnetic resonance imaging/ 879332

32 (MRI or magnetic resonance or mr imaging or MDCT).tw. 1391223
33 contrast-enhanced ultrasound/ 5671
34 (CEUS or contrast enhanced ultrasound).tw. 14486
35 or/29-34 3207013
36 28 and 35 1405
37 36 use emczd 858
38 limit 37 to dc=20190912-20220113 456
39 (LI-RADS or LIRADS).tw,kw. 1194
40 "liver imaging reporting and data system".tw,kw. 820
41 39 or 40 1273
42 (LR-1 or LR-2 or LR-3 or LR-4 or LR-5 or LR-5V or LR-OM or LR-TIV or LR-M or LR1 or LR2 or LR3 or LR4 or LR5 or LR5V or LRTIV or LRM or LROM).tw. 5129
43 Radiology Information Systems/ 6755
44 42 or 43 11848
45 Carcinoma, Hepatocellular/ or Cholangiocarcinoma/208970
46 (hepatocellular carcinoma* or hepatocellular neoplasm* or hepatocellular cancer or hepatic cell carcinoma* or HCC or cholangiocarcinoma* or hepatic nodule* or liver lesion* or adrenocortical carcinoma*).tw,kw. 337332
47 Liver Neoplasms/ 186988
48 exp Liver Diseases/ or liver disease*.tw,kw. 1838214
49 Liver/ 1023574
50 liver imaging.tw,kw. 3255
51 or/45-50 2670843
52 44 and 51 941
53 41 or 52 1663
54 53 use cctr 41
55 limit 54 to yr="2019 -Current" 12
56 24 or 38 or 55 770

Scopus – Jan 13, 2022
Search Strategy:
442 References
((TITLE-ABS-KEY (li-rads OR lirads)) OR (TITLE-ABS-KEY ("liver imaging reporting and data system")) OR (TITLE-ABS-KEY (lr-1 OR lr-2 OR lr-3 OR lr-4 OR lr-5 OR lr-5v OR lr-om OR lr-tiv OR lr-m OR lr1 OR lr2 OR lr3 OR lr4 OR lr5 OR lr5v OR lrtiv OR lrm OR lrom))) AND ((TITLE-ABS-KEY ("computed tomograph*" OR ct)) OR (TITLE-ABS-KEY (mri OR magnetic AND resonance OR mdct)) OR (TITLE-ABS-KEY (ceus OR contrast AND enhanced AND ultrasound))) AND (LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019))

Appendix 3.3. Data dictionary send to authors contributing individual participant data (IPD). Data was originally collected for the previous IPD study (13). This shortened list of questions were extracted from the original data and used for the purpose of this review.

Definition	**Values**
Dataset identifier assigned to dataset by Registry team to each dataset	00001, 00002, 00003…
Primary observation identifier used in primary database	As recorded in primary database
Primary patient identifier used in primary database	As recorded in primary database
Patient's age, measured continuously	As recorded in primary database
Patient's gender	0 = Female 1 = Male
Country where patient recruited.	As recorded in primary database
What was the reference standard	0 = Biopsy 1 = Resection pathology 2 = Explant pathology 3 = Other (specify)
If the reference in the previous column was recorded as "other," please specify.	As recorded in primary database
Maximum diameter of the observation (mm)	As recorded in primary database
final diagnosis	0 = None 1 = HCC 2 = Non-HCC malignancy (specify) 3 = Benign 4 = Other (specify)
If the final diagnosis in the previous column was recorded as "other," please specify.	As recorded in primary database
Histological differentiation	0 = Not applicable (not a tumor) 1 = Very well 2 = Well 3 = Moderately 4 = Poorly
Version of LI-RADS used	1 = CT/MR v2018 2 = CT/MR v2017 3 = CT/MR v2014 4 = CT/MR v2013 5 = CEUS v2017 6 = CEUS v2016
Imaging Modality used	1 = CT 2 = MRI 3 = CT + MRI 4 = CEUS
CT Data	
LI-RADS data	
LI-RADS category	0 = NC 1 = 1 2 = 2 3 = 3 4 = 4 5 = 5 6 = M

	7 = TIV
How was LI-RADS category assigned	0 = Prospective clinical radiology report 1 = Retrospective research read of clinical exam 2 = Other (specify)
If the LI-RADS category assigned in the previous column was recorded as "other," please specify.	As recorded in primary database
Technical Parameters	
Use of a delayed phase on CT	0 = without 3-minute delayed phase 1 = > 3-minute delayed phase
Use of unenhanced phase on CT	0 = without unenhanced phase 1 = with unenhanced phase
Number of detectors on CT	As recorded in primary database
CT contrast generic name	As recorded in primary database
LI-RADS Major Features	
CT Size (mm)	0 = <10 mm 1 = 10-19 mm 2 = ≥20 mm
CT nonrim Arterial phase hyperenhancement (APHE)	0 = No 1 = Yes
CT Nonperipheral "washout"	0 = No 1 = Yes
CT Enhancing "capsule"	0 = No 1 = Yes
CT Threshold growth LIRADS v2014-2017	0 = No 1 = Yes
CT Threshold growth LIRADS v2018	0 = No 1 = Yes
Ancillary Features	
US visibility as discrete nodule	0 = Absent 1 = Present 2= Not evaluated 99= Not applicable (Ultrasound not conducted)
Subthreshold growth	0 = Absent 1 = Present 2= Not evaluated 99= Not applicable (No previous measurement of the mass)
Corona enhancement	0 = Absent 1 = Present 2= Not evaluated
Fat sparing in solid mass	0 = Absent 1 = Present 2= Not evaluated
Nonenhancing "capsule"	0 = Absent 1 = Present 2= Not evaluated
Nodule-in-nodule	0 = Absent 1 = Present

	2= Not evaluated
Mosaic architecture	0 = Absent 1 = Present 2= Not evaluated
Blood products in mass	0 = Absent 1 = Present 2= Not evaluated
Fat in mass, more than adjacent liver	0 = Absent 1 = Present 2= Not evaluated
Size stability > 2 years	0 = Absent 1 = Present 2= Not evaluated 99= Not applicable (No previous measurement of the mass)
Size reduction	0 = Absent 1 = Present 2= Not evaluated 99= Not applicable (No previous measurement of the mass)
Parallels blood pool	0 = Absent 1 = Present 2= Not evaluated
Undistorted vessels	0 = Absent 1 = Present 2= Not evaluated
Iron in mass, more than liver	0 = Absent 1 = Present 2= Not evaluated
MRI Data	
LI-RADS data	
LI-RADS category	0 = NC 1 = 1 2 = 2 3 = 3 4 = 4 5 = 5 6 = M 7 = TIV
How was LI-RADS category assigned	0 = Prospective clinical radiology report 1 = Retrospective research read of clinical exam 2 = Other (specify)
If the LI-RADS category assigned in the previous column was recorded as “other,” please specify.	As recorded in primary database
Technical Parameters	
Type of agent used	0 = use of extracellular agent 1 = gadobenate without HBP 2 = gadobenate with HBP 3 = gadoxetate disodium

MRI contrast generic name	As recorded in primary database
Amount administered of contrast agent	As recorded in primary database
Dilution of contrast agent	0 = Not diluted 1 = Diluted
MRI field strength	0 = 1.5 T 1 = 3.0 T
LI-RADS Major Features	
MRI Size (mm)	0 = <10 mm 1 = 10-19 mm 2 = ≥20 mm
MRI nonrim Arterial phase hyperenhancement (APHE)	0 = No 1 = Yes
MRI Nonperipheral “washout”	0 = No 1 = Yes
MRI Enhancing “capsule”	0 = No 1 = Yes
MRI Threshold growth LIRADS v2014-2017	0 = No 1 = Yes
MRI Threshold growth LIRADS v2018	0 = No 1 = Yes
Ancillary Features	
US visibility as discrete nodule	0 = Absent 1 = Present 2= Not evaluated 99= Not applicable (Ultrasound not conducted)
Subthreshold growth	0 = Absent 1 = Present 2= Not evaluated 99= Not applicable (No previous measurement of the mass)
Restricted diffusion	0 = Absent 1 = Present 2= Not evaluated
Mild-moderate T2 hyperintensity	0 = Absent 1 = Present 2= Not evaluated
Corona enhancement	0 = Absent 1 = Present 2= Not evaluated
Fat sparing in solid mass	0 = Absent 1 = Present 2= Not evaluated
Iron sparing in solid mass	0 = Absent 1 = Present 2= Not evaluated
Transitional phase hypointensity	0 = Absent 1 = Present 2= Not evaluated

	99= Not applicable (Hepatobiliary agent not used)
Hepatobiliary phase hypointensity	0 = Absent 1 = Present 2= Not evaluated 99= Not applicable (Hepatobiliary agent not used)
Nonenhancing “capsule”	0 = Absent 1 = Present 2= Not evaluated
Nodule-in-nodule	0 = Absent 1 = Present 2= Not evaluated
Mosaic architecture	0 = Absent 1 = Present 2= Not evaluated
Blood products in mass	0 = Absent 1 = Present 2= Not evaluated
Fat in mass, more than adjacent liver	0 = Absent 1 = Present 2= Not evaluated
Size stability > 2 years	0 = Absent 1 = Present 2= Not evaluated 99= Not applicable (No previous measurement of the mass)
Size reduction	0 = Absent 1 = Present 2= Not evaluated 99= Not applicable (No previous measurement of the mass)
Parallels blood pool	0 = Absent 1 = Present 2= Not evaluated
Undistorted vessels	0 = Absent 1 = Present 2= Not evaluated
Iron in mass, more than liver	0 = Absent 1 = Present 2= Not evaluated
Marked T2 hyperintensity	0 = Absent 1 = Present 2= Not evaluated
Hepatobiliary phase iso-intensity	0 = Absent 1 = Present 2= Not evaluated 99= Not applicable (Hepatobiliary agent not used)
CEUS Data	

LI-RADS data	
LI-RADS category	0 = NC 1 = 1 2 = 2 3 = 3 4 = 4 5 = 5 6 = M 7 = TIV
How was LI-RADS category assigned	0 = Prospective clinical radiology report 1 = Retrospective research read of clinical exam 2 = Other (specify)
If the LI-RADS category assigned in the previous column was recorded as "other," please specify.	As recorded in primary database
LI-RADS Major Features	
CEUS Size (mm)	0 = <10 mm 1 = 10-19 mm 2 = ≥20 mm
CEUS Arterial phase hyperenhancement (APHE, not rim, not peripheral discontinuous)	0 = No 1 = not rim / not peripheral discontinuous globular 2 = rim or peripheral discontinuous globular
CEUS Nonperipheral "washout"	0 = no washout of any type 1 = late and mild washout 2 = early (< 60s) washout 3= marked washout
Ancillary Features	
Definite growth	0 = Absent 1 = Present 2= Not evaluated 99= Not applicable (No previous measurement of the mass)
Nodule-in-nodule architecture	0 = Absent 1 = Present 2= Not evaluated
Mosaic architecture	0 = Absent 1 = Present 2= Not evaluated
CEUS Size stability ≥ 2 years	0 = Absent 1 = Present 2= Not evaluated 99= Not applicable (No previous measurement of the mass)
CEUS Size reduction	0 = Absent 1 = Present 2= Not evaluated 99= Not applicable (No previous measurement of the mass)

Appendix 3.4. Modified QUADAS-2 signalling questions to determine the risk of bias and applicability concerns of included studies. Signalling questions were applied from the previous individual participant data (IPD) study (13).

Domain 1: Patient and observation selection	
C. Risk of bias	
Describe methods of patient and observation selection:	
I. Was a consecutive sample of patients, random sample of patients, or all patients over a given time period enrolled?	Yes, no or unclear
II. Was a case-control design avoided?	Yes, no or unclear
III. Did the study avoid inappropriate exclusions? (e.g. If a non-pathology-based reference standard was used, might it have inappropriately excluded observations?)	Yes, no or unclear
IV. Could the selection of patients have introduced bias? (Low= "yes" to all, otherwise refer to phase 2)	Yes, no or unclear
Could the selection of patients have introduced bias?	Risk: low, high or unclear
D. Concerns regarding applicability	
Describe included patients and included observations (prior testing, presentation, intended use of multiphase CT/MRI and setting):	
Is there concern that the included patients do not match the review question?	Concern: low, high or unclear
Domain 2: Index test	
CT	
A. Risk of bias	
Describe how multiphase contrast-enhanced CT was conducted and interpreted	
I. Were the multiphase CT results interpreted without knowledge of the results of the reference standard?	Yes, no or unclear
II. Were the multiphase CT results unlikely to be biased by findings on other imaging exams?	Yes, no or unclear
III. Was a delayed phase consistently used? (delayed between 3-5 mins has increased sensitivity for washout)	Yes, no or unclear
IV. Was the index test interpreted by more than one radiologist and were discrepancies resolved in an objective way?	Yes, no or unclear
Could the conduct or interpretation of the multiphase CT have introduced bias? (Low= "yes" to all, otherwise refer to phase 2)	Risk: low, high or unclear
B. Concerns regarding applicability	
Is there concern that the multiphase CT, its conduct, or interpretation differ from the review question?	Concern: low, high or unclear
MRI	
A. Risk of bias	
Describe how multiphase contrast-enhanced MRI was conducted and interpreted	
I. Were the MRI results interpreted without knowledge of the results of the reference standard?	Yes, no or unclear
II. Were the MRI results unlikely to be biased by findings on other imaging exams?	Yes, no or unclear

III. Was the index test interpreted by more than one radiologist and were discrepancies resolved in an objective way?	Yes, no or unclear
Could the conduct or interpretation of the MRI have introduced bias? (Low= "yes" to all, otherwise refer to phase 2)	Risk: low, high or unclear
B. Concerns regarding applicability Is there concern that the MRI, its conduct, or interpretation differ from the review question?	Concern: low, high or unclear
<u>CEUS</u> A. Risk of bias *Describe how contrast-enhanced US (CEUS) was conducted and interpreted*	
I. Were the CEUS results interpreted without knowledge of the results of the reference standard?	Yes, no or unclear
II. Were the CEUS results unlikely to be biased by findings on other imaging exams?	Yes, no or unclear
III. Did CEUS technical parameters meet the minimum standard in LI-RADS?	Yes, no or unclear
IV. Was the index test interpreted by more than one radiologist and were discrepancies resolved in an objective way?	Yes, no or unclear
Could the conduct or interpretation of the CEUS have introduced bias? (Low= "yes" to all, otherwise refer to phase 2)	Risk: low, high or unclear
B. Concerns regarding applicability Is there concern that the CEUS, its conduct, or interpretation differ from the review question?	Concern: low, high or unclear
<u>Domain 3: Reference standard</u> C. Risk of bias *Describe the reference standard and how it was conducted and interpreted*	
I. Is the reference standard likely to correctly classify the target condition at the observation level (particularly non-pathology-based reference standards and also explant reference standards – was the method of lesion matching described and likely to be robust)?	Yes, no or unclear
II. Were the reference standard results interpreted without knowledge of the results of the CT, MRI or CEUS?	Yes, no or unclear
Could the reference standard, its conduct or interpretation have introduced bias? (Low= "yes" to all, otherwise refer to phase 2)	Risk: low, high or unclear
D. Concerns regarding applicability Is there concern that the target condition as defined by the reference standard does not match the review question?	Concern: low, high or unclear
<u>Domain 4: Flow and timing</u> A. Risk of bias *Describe any patients who did not receive multiphase CT, MRI, CEUS, and/or reference standard, or who were excluded from the 2 x 2 table (refer to flow diagram)*	

Describe the time interval and any interventions between multiphase CT, MRI, CEUS and reference standard *Describe any observations within included patients who were excluded from the 2 x 2 table*	Yes, no or unclear
I. Was there an appropriate interval between multiphase CT, MRI or CEUS and reference standard?	Yes, no or unclear
II. Did all patients and observations receive a reference standard?	Yes, no or unclear
III. Did patients and observations receive the same reference standard?	Yes, no or unclear
IV. Were all patients and observations included in the analysis?	Yes, no or unclear
V. Is there unlikely to be a selection bias from selection of patients only with liver explanation	Yes, no or unclear
VI. Is there unlikely to be verification bias from tissue sampling of only a subset of observations?	
VII. Is there unlikely to be incorporation bias for studies including LI-RADS 5 as a reference standard?	Yes, no or unclear
Could the patient and observation flow have introduced bias? (Low= "yes" to all, otherwise refer to phase 2)	Risk: low, high or unclear

Appendix 3.5. Reporting of CEUS ancillary features in included studies.

Ancillary Feature	Absent	Present	Not reported/ Not applicable	Percent reported	Studies reporting AF
Definite growth	0	0	185	28.0	
Nodule-in-nodule architecture	164	21	0	100.0	Hu, Makoyeva
Mosaic architecture	168	17	0	100.0	Hu, Makoyeva
CEUS size stability ≥2 years	0	0	185	28.0	
CEUS size reduction	0	0	185	28.0	

Abbreviations- CEUS: contrast enhanced ultrasound

Appendix 3.6. Multivariable analysis of each ancillary feature, adjusted for major features (except threshold growth), to determine the odds of HCC.

Ancillary Feature	Association with (Based on LIRADS v2018)	OR (95% CI)	Enhancing capsule (95% CI)	APHE (95% CI)	Size 10-19 mm (95% CI)	Size ≥ 20mm (95% CI)	Nonperipheral washout (95% CI)
US visibility as discrete nodule	Malignancy	**3.18 (1.77-5.73)***	**6.39 (2.72-15.04)***	**8.54 (5.60-13.03)***	1.23 (0.65-2.33)	**3.63 (1.82-7.23)***	**3.24 (2.16-4.85)***
Subthreshold growth	Malignancy	**3.21 (1.26-8.17)***	**3.65 (1.24-10.78)***	**3.15 (1.39-7.16)***	3.02 (0.88-10.35)	**3.65 (1.01-13.23)***	**3.66 (1.68-7.94)***
Restricted diffusion	Malignancy	**3.18 (2.52-4.02)***	**2.57 (1.95-3.40)***	**3.66 (2.87-4.66)***	**1.49 (1.07-2.07)***	**1.64 (1.17-2.28)***	**3.36 (2.67-4.24)***
Mild-moderate T2 hyperintensity	Malignancy	**3.05 (2.39-3.89)***	**2.65 (2.02-3.49)***	**3.98 (3.13-5.06)***	**1.64 (1.18-2.28)***	**1.68 (1.20-2.35)***	**3.34 (2.67-4.19)***
Corona enhancement	Malignancy	1.18 (0.75-1.87)	**3.49 (2.57-4.74)***	**5.76 (4.48-7.41)***	**1.69 (1.22-2.35)***	**2.52 (1.81-3.51)***	**4.93 (3.90-6.24)***
Fat sparing in solid mass	Malignancy	0.81 (0.44-1.49)	**3.48 (2.57-4.73)***	**5.92 (4.60-7.63)***	**1.57 (1.12-2.20)***	**2.37 (1.69-3.31)***	**4.89 (3.85-6.21)***
Iron sparing in solid mass	Malignancy	**2.63 (1.30-5.31)***	**3.17 (2.30-4.36)***	**6.03 (4.57-7.97)***	**1.95 (1.32-2.88)***	**2.86 (1.94-4.20)***	**4.77 (3.67-6.19)***
Transitional phase hypointensity	Malignancy	**2.08 (1.48-2.93)***	**1.79 (1.28-2.49)***	**4.85 (3.69-6.38)***	**1.91 (1.26-2.88)***	**1.88 (1.25-2.81)***	**2.28 (1.70-3.07)***
Hepatobiliary phase hypointensity	Malignancy	**2.06 (1.42-3.00)***	**1.93 (1.39-2.68)***	**4.64 (3.54-6.08)***	**2.00 (1.33-3.01)***	**2.01 (1.35-3.00)***	**2.72 (2.08-3.56)***
Growth (threshold + subthreshold)	Malignancy	**2.68 (1.42-5.07)***	**2.84 (1.33-6.08)***	**4.72 (2.55-8.73)***	1.83 (0.90-3.72)	**2.52 (1.23-5.17)***	**3.56 (1.95-6.50)***
Nonenhancing "capsule"	HCC	**3.50 (1.53-8.01)***	**3.49 (2.44-4.97)***	**6.73 (5.02-9.00)***	**1.84 (1.27-2.68)***	**2.49 (1.71-3.60)***	**5.39 (4.08-7.12)***
Nodule-in-nodule	HCC	**2.15 (1.15-4.04)***	**2.65 (2.01-3.49)***	**4.32 (3.39-5.49)***	**2.04 (1.47-2.82)***	**2.34 (1.70-3.21)***	**4.03 (3.20-5.07)***
Mosaic architecture	HCC	**2.53 (1.58-4.04)***	**3.14 (2.30-4.28)***	**5.52 (4.29-7.09)***	**1.72 (1.24-2.39)***	**2.32 (1.66-3.23)***	**4.82 (3.80-6.10)***
Blood products in mass	HCC	1.28 (0.83-1.96)	**2.83 (2.15-3.72)***	**3.92 (3.07-5.01)***	**2.07 (1.47-2.91)***	**2.88 (2.06-4.03)***	**3.97 (3.15-4.99)***
Fat in mass, more than adjacent liver	HCC	**1.44 (1.03-2.01)***	**2.85 (2.18-3.73)***	**4.15 (3.29-5.23)***	**1.75 (1.28-2.39)***	**2.41 (1.77-3.28)***	**3.80 (3.05-4.73)***
Size stability > 2 years	Benignity	0.45 (0.13-1.55)	**3.62 (1.29-10.16)***	**4.84 (2.14-10.96)***	4.12 (0.97-17.54)	3.33 (0.78-14.23)	**4.09 (1.88-8.87)***
Size reduction	Benignity	0.28 (0.01-12.34)	**3.39 (1.24-9.30)***	**3.74 (1.74-8.06)***	**5.43 (1.78-16.54)***	**5.91 (1.89-18.55)***	**5.87 (2.92-11.82)***
Parallels blood pool	Benignity	**0.07 (0.01-0.49)***	**2.40 (1.66-3.47)***	**7.06 (5.13-9.73)***	**2.79 (1.85-4.22)***	**4.61 (3.04-6.97)***	**5.81 (4.26-7.92)***
Undistorted vessels	Benignity	0.90 (0.30-2.68)	**2.45 (1.70-3.53)***	**6.85 (5.00-9.40)***	**2.83 (1.88-4.26)***	**4.63 (3.08-6.98)***	**5.98 (4.41-8.12)***
Iron in mass, more than liver	Benignity	0.27 (0.07-1.11)	**2.17 (1.49-3.15)***	**6.92 (5.01-9.56)***	**2.36 (1.54-3.63)***	**3.94 (2.59-5.99)***	**5.77 (4.19-7.95)***
Marked T2 hyperintensity	Benignity	**0.18 (0.07-0.45)***	**2.33 (1.62-3.36)***	**7.69 (5.67-10.43)***	**2.52 (1.72-3.69)***	**4.08 (2.75-6.03)***	**5.09 (3.80-6.83)***
Hepatobiliary phase iso-intensity	Benignity	**0.46 (0.27-0.80)***	**2.68 (1.81-3.99)***	**8.11 (6.01-10.95)***	**1.89 (1.22-2.92)***	**1.92 (1.24-2.96)***	**4.06 (3.03-5.43)***

*Significant association ($p<0.05$); Abbreviations- APHE: arterial phase hyperenhancement; HCC: hepatocellular carcinoma; US: ultrasound; CI: confidence interval

Appendix 3.7. Multivariable analysis of each ancillary feature, adjusted for major features (except threshold growth), to determine the odds of malignancy.

Ancillary Feature	Association with (Based on LIRADS v2018)	OR (95% CI)	Enhancing capsule (95% CI)	APHE (95% CI)	Size 10-19 mm (95% CI)	Size ≥ 20mm (95% CI)	Nonperipheral washout (95% CI)
US visibility as discrete nodule	Malignancy	2.09 (0.83-5.26)	**16.04 (2.04-125.93)***	**14.62 (8.32-25.71)***	**2.13 (1.03-4.43)***	**12.71 (5.37-30.12)***	**2.89 (1.66-5.02)***
Subthreshold growth	Malignancy	**5.03 (1.64-15.46)***	**4.79 (1.23-18.57)***	2.07 (0.89-4.84)	2.99 (0.92-9.72)	**8.16 (2.29-29.02)***	**2.88 (1.27-6.50)***
Restricted diffusion	Malignancy	**14.45 (9.82-21.27)***	**2.56 (1.78-3.70)***	**3.81 (2.20-6.60)***	1.51 (0.98-2.32)	**5.30 (3.33-8.43)***	**2.21 (1.51-3.22)***
Mild-moderate T2 hyperintensity	Malignancy	**10.18 (7.17-14.44)***	**3.57 (2.10-6.06)***	**2.88 (2.02-4.12)***	**1.65 (1.08-2.51)***	**4.84 (3.06-7.65)***	**2.50 (1.74-3.58)***
Corona enhancement	Malignancy	**4.89 (1.71-14.01)***	**4.07 (2.45-6.76)***	**3.52 (2.51-4.93)***	**1.67 (1.12-2.49)***	**6.69 (4.34-10.31)***	**3.46 (2.46-4.86)***
Fat sparing in solid mass	Malignancy	0.79 (0.30-2.09)	**4.38 (2.60-7.39)***	**3.73 (2.62-5.32)***	1.50 (0.98-2.30)	**6.61 (4.17-10.47)***	**3.31 (2.30-4.75)***
Iron sparing in solid mass	Malignancy	2.69 (0.52-13.92)	**3.61 (1.90-6.88)***	**4.42 (2.78-7.02)***	**2.09 (1.18-3.68)***	**9.49 (5.20-17.32)***	**3.78 (2.33-6.14)***
Transitional phase hypointensity	Malignancy	**2.28 (1.45-3.59)***	**3.22 (1.52-6.81)***	**8.32 (5.62-12.31)***	**2.68 (1.68-4.30)***	**25.35 (14.54-44.22)***	**2.10 (1.35-3.28)***
Hepatobiliary phase hypointensity	Malignancy	**3.15 (1.95-5.07)***	**3.11 (1.59-6.08)***	**7.46 (5.11-10.90)***	**2.79 (1.76-4.45)***	**23.88 (13.86-41.13)***	**2.40 (1.63-3.55)***
Growth (threshold + subthreshold)	Malignancy	**4.36 (1.87-10.17)***	**6.62 (1.83-10.17)***	1.00 (0.50-2.01)	**0.38 (0.17-0.83)***	**0.66 (0.27-1.58)**	**2.63 (1.29-5.39)***
Nonenhancing "capsule"	HCC	2.06 (0.61-6.91)	**4.74 (2.56-8.78)***	**4.34 (2.91-6.48)***	1.44 (0.91-2.29)	**8.43 (5.00-14.21)***	**3.31 (2.20-4.98)***
Nodule-in-nodule	HCC	1.43 (0.46-4.43)	**4.13 (2.48-6.87)***	**3.34 (2.38-4.69)***	**1.77 (1.18-2.64)***	**8.19 (5.29-12.68)***	**3.41 (2.42-4.80)***
Mosaic architecture	HCC	**3.26 (1.25-8.51)***	**3.89 (2.32-6.52)***	**3.35 (2.38-4.71)***	**1.71 (1.14-2.56)***	**6.78 (4.36-10.54)***	**3.37 (2.39-4.76)***
Blood products in mass	HCC	2.31 (0.81-6.57)	**3.62 (2.04-6.45)***	**3.22 (2.15-4.83)***	**2.12 (1.31-3.45)***	**9.74 (5.79-16.38)***	**4.32 (2.85-6.54)***
Fat in mass, more than adjacent liver	HCC	0.71 (0.42-1.19)	**4.25 (2.54-7.08)***	**3.42 (2.43-4.82)***	**1.77 (1.18-2.66)***	**8.26 (5.32-12.83)***	**3.59 (2.53-5.10)***
Size stability > 2 years	Benignity	0.63 (0.18-2.14)	**4.45 (1.21-16.34)***	**3.30 (1.45-7.54)***	**4.70 (1.25-17.69)***	**8.18 (2.09-32.09)***	**3.33 (1.46-7.62)***
Size reduction	Benignity	0.17 (0.003-10.75)	**3.81 (1.10-13.21)***	**2.54 (1.12-5.75)***	**6.07 (1.98-18.55)***	**13.50 (4.17-43.77)***	**4.88 (2.29-10.40)***
Parallels blood pool	Benignity	**0.02 (0.002-0.24)***	**4.32 (2.33-8.00)***	**4.08 (2.62-6.34)***	**2.02 (1.21-3.38)***	**7.98 (4.53-14.05)***	**4.67 (2.98-7.32)***
Undistorted vessels	Benignity	0.72 (0.17-3.12)	**4.40 (2.44-7.94)***	**3.71 (2.43-5.64)***	**2.01 (1.22-3.29)***	**7.40 (4.32-12.68)***	**4.77 (3.11-7.31)***
Iron in mass, more than liver	Benignity	**0.07 (0.01-0.44)***	**4.66 (2.42-8.98)***	**4.10 (2.56-6.57)***	1.56 (0.88-2.76)	**5.84 (3.18-10.73)***	**4.39 (2.68-7.19)***
Marked T2 hyperintensity	Benignity	**0.06 (0.02-0.19)***	**4.54 (2.61-7.91)***	**4.41 (3.02-6.43)***	**1.87 (1.21-2.91)***	**7.19 (4.39-11.77)***	**3.46 (2.37-5.06)***
Hepatobiliary phase iso-intensity	Benignity	**0.16 (0.08-0.30)***	**3.22 (1.63-6.36)***	**8.97 (6.05-13.30)***	**2.69 (1.67-4.33)***	**17.95 (10.40-30.99)***	**2.65 (1.79-3.92)***

*Significant association ($p<0.05$); Abbreviations- APHE: arterial phase hyperenhancement; HCC: hepatocellular carcinoma; US: ultrasound; CI: confidence interval

Appendix 3.8. Results from the secondary analysis

The presence or absence of TG was reported in four studies, with an average reporting rate of 21.5% (Range: 5.8%-73.8%) for each AF (Appendix 3.9). The results of the secondary analysis (with TG included as a major feature) are provided in (Appendices 3.9-3.11). AFs were assessed independently with the number of observations, patients and the studies included in each model indicated in appendix 3.9. The heterogeneity within each random effects model is presented in appendix 3.12. All models had highly variable I^2 values due to the limited sample size.

Transitional phase hypointensity and hepatobiliary phase hypointensity could not be evaluated due to a lack of studies. Of the remaining seven AFs favoring malignancy in general, only subthreshold growth (OR: 11.92 [95% CI=1.43-99.60]), restricted diffusion (OR: 4.03 [95% CI=1.42-11.46]) and mild-moderate T2 hyperintensity (OR: 3.42 [95% CI=1.44-8.13]) were associated with increased odds of malignancy. Subthreshold growth (OR: 8.85 [95% CI=1.78-43.99]) and iron sparing in solid mass (OR: 5.77 [95% CI=1.05-31.80]) were significantly associated with increased odds of HCC.

Of the five AFs favoring HCC in particular, none were independently associated with HCC or malignancy. Hepatobiliary phase iso-intensity and parallels blood pool could not be evaluated due to a lack of studies and positive observations respectively. Of the remaining five AFs favoring benignity, none showed a significant association with benignity, HCC or malignancy.

In the secondary analysis the average positivity rates were as follows: 27.0% (range, 1.6%-59.6%) for AFs favoring malignancy in general; 9.2% (range, 1.3%-15.9%) for AFs favoring HCC in particular; and 4.5% (range, 0%-11.8%) for AFs favoring benignity.

Appendix 3.9. Multivariable analysis of each ancillary feature, adjusted for major features (APHE, size, nonperipheral washout, enhancing capsule and threshold growth) to determine the odds of HCC and malignancy.

Feature	Association with (Based on LIRADS v2018)	Observations (Patients, Studies)	HCC observations n(%)	Malignant observations n(%)	Positive Data points n(%)	Study sources	HCC		Malignancy	
							% of patients reporting threshold growth	OR (95%CI)	% of patients reporting threshold growth	OR (95%CI)
US visibility as discrete nodule	Malignancy	52 (52,2)	39 (75.0)	45 (86.5)	31 (59.6)	(22), (39)	5.8	1.83 (0.31-10.60)	6.4	0.95 (0.13-7.10)
Subthreshold growth	Malignancy	184 (155,3)	134 (72.8)	147 (79.9)	46 (25.0)	(27), (22), (39)	68.9	**8.85 (1.78-43.99)***	68.9	**11.92 (1.43-99.60)***
Restricted diffusion	Malignancy	285 (251,4)	183 (64.2)	243 (85.3)	106 (37.2)	(27), (22), (23), (39)	10.1	0.99 (0.50-1.98)	10.5	**4.03 (1.42-11.46)***
Mild-moderate T2 hyperintensity	Malignancy	308 (261,4)	199 (64.6)	260 (84.4)	156 (50.7)	(27), (22), (23), (39)	10.5	1.07 (0.58-1.99)	10.8	**3.42 (1.44-8.13)***
Corona enhancement	Malignancy	312 (265,4)	201 (64.4)	264 (84.6)	5 (1.6)	(27), (22), (23), (39)	10.6	0.65 (0.05-8.17)	11.0	8.52e+06 (0.00-Inf)
Fat sparing in solid mass	Malignancy	234 (218,4)	141 (60.3)	201 (85.9)	14 (6.0)	(27), (22), (23), (39)	8.4	0.85 (0.23-3.16)	8.7	0.47 (0.08-2.85)
Iron sparing in solid mass	Malignancy	208 (201,4)	126 (60.6)	186 (89.4)	19 (9.1)	(27), (22), (23), (39)	8.4	**5.77 (1.05-31.80)***	8.8	1.86 (0.34-10.28)
Transitional phase hypointensity	Malignancy	--- (---,0)	---	---	---	---	---	---	---	---
Hepatobiliary phase hypointensity	Malignancy	--- (---,0)	---	---	---	---	---	---	---	---
Nonenhancing "capsule"	HCC	102 (102,2)	51 (50.0)	100 (98.0)	12 (1.8)	(23), (39)	4.2	2.59 (0.30-22.39)	4.2	1.57e-32 (0.00-Inf)
Nodule-in-nodule	HCC	312 (265,4)	201 (64.4)	264 (84.6)	4 (1.3)	(27), (22), (23), (39)	10.5	1.87e+06 (0.00-Inf)	10.5	3.35e+06 (0.00-Inf)
Mosaic architecture	HCC	312 (265,4)	201 (64.4)	264 (84.6)	28 (9.0)	(27), (22), (23), (39)	10.6	2.48 (0.49-12.49)	11.0	1.90e+07 (0.00-Inf)
Blood products in mass	HCC	308 (261,4)	199 (64.6)	260 (84.4)	25 (8.1)	(27), (22), (23), (39)	10.7	1.13 (0.36-3.53)	11.1	0.83 (0.17-4.05)
Fat in mass, more than adjacent liver	HCC	308 (261,4)	199 (64.6)	260 (84.4)	49 (15.91)	(27), (22), (23), (39)	10.1	1.05 (0.42-2.63)	10.4	0.47 (0.16-1.41)
Size stability > 2 years	Benignity	186 (161,3)	139 (74.7)	152 (81.7)	22 (11.8)	(27), (22), (39)	73.8	0.35 (0.10-1.24)	73.8	0.45 (0.12-1.67)
Size reduction	Benignity	216 (177,3)	159 (73.6)	173 (80.1)	3 (1.4)	(27), (22), (39)	62.1	0.24 (0.01-4.40)	62.1	0.15 (0.01-3.96)

Parallels blood pool	Benignity	312 (265,4)	201 (64.4)	264 (84.6)	0 (0)	(27), (22), (23), (39)	12.7	**No positive observations**	12.7	**No positive observations**
Undistorted vessels	Benignity	312 (265,4)	201 (64.4)	264 (84.6)	20 (6.4)	(27), (22), (23), (39)	12.7	1.20 (0.29-4.97)	12.7	1.31 (0.20-8.75)
Iron in mass, more than liver	Benignity	193 (193,3)	114 (59.1)	173 (89.6)	1 (0.5)	(22), (23), (39)	9.3	3.07e-07 (0.00-Inf)	9.3	7.39e-09 (0.00-Inf)
Marked T2 hyperintensity	Benignity	308 (261,4)	199 (64.6)	260 (84.4)	22 (7.1)	(27), (22), (23), (39)	12.4	0.49 (0.15-1.62)	12.4	0.34 (0.08-1.41)
Hepatobiliary phase iso-intensity	Benignity	--- (---,0)	---	---	---	---	---	---	---	---

*Significant association ($p<0.05$); Abbreviations- APHE: arterial phase hyperenhancement; HCC: hepatocellular carcinoma; US: ultrasound; CI: confidence interval

Appendix 3.10. Multivariable analysis of each ancillary feature, adjusted for all major features (APHE, size, nonperipheral washout, enhancing capsule and threshold growth), to determine the odds of HCC.

Ancillary Feature	Association with (Based on LIRADS v2018)	OR (95% CI)	Enhancing capsule (95% CI)	APHE (95% CI)	Size 10-19 mm (95% CI)	Size ≥ 20mm (95% CI)	Nonperipheral washout (95% CI)	Threshold Growth (95% CI)
US visibility as discrete nodule	Malignancy	1.83 (0.31-10.60)	5.82 (0.70-48.42)	**5.95 (1.08-32.69)***	5.48e-07 (0.00-Inf)	8.22e-08 (0.00-Inf)	**9.80 (1.35-71.19)***	0.79 (0.10-6.11)
Subthreshold growth	Malignancy	**8.85 (1.78-43.99)***	2.85 (0.93-8.75)	**2.97 (1.20-7.39)***	2.71 (0.68-10.75)	1.72 (0.40-7.37)	**4.03 (1.50-10.84)***	0.86 (0.26-2.82)
Restricted diffusion	Malignancy	0.99 (0.50-1.98)	2.15 (0.96-4.78)	**5.86 (2.98-11.52)***	1.92 (0.87-4.25)	**2.26 (1.04-4.88)***	**3.29 (1.57-6.89)***	1.46 (0.65-3.32)
Mild-moderate T2 hyperintensity	Malignancy	1.07 (0.58-1.99)	**2.35 (1.07-5.17)***	**6.02 (3.12-11.61)***	**2.42 (1.13-5.18)***	**2.58 (1.22-5.46)***	**3.26 (1.62-6.58)***	1.59 (0.74-3.44)
Corona enhancement	Malignancy	0.65 (0.05-8.17)	**2.33 (1.07-5.11)***	**5.98 (3.13-11.43)***	**2.40 (1.12-5.34)***	**2.54 (1.21-5.34)***	**3.43 (1.71-6.88)***	1.59 (0.73-3.43)
Fat sparing in solid mass	Malignancy	0.85 (0.23-3.16)	2.20 (0.97-4.97)	**6.61 (3.24-13.47)***	2.41 (0.98-5.92)	2.07 (0.94-4.56)	**2.73 (1.23-6.08)***	1.34 (0.57-3.18)
Iron sparing in solid mass	Malignancy	**5.77 (1.05-31.80)***	2.20 (0.96-5.06)	**7.22 (3.41-15.29)***	1.54 (0.56-4.24)	**1.28 (0.55-2.97)**	**2.51 (1.07-5.93)***	1.18 (0.47-2.93)
Transitional phase hypointensity	Malignancy	---	---	---	---	---	---	---
Hepatobiliary phase hypointensity	Malignancy	---	---	---	---	---	---	---
Nonenhancing "capsule"	HCC	2.59 (0.30-22.39)	1.05 (0.26-4.31)	**23.35 (6.13-88.96)***	0.26 (0.03-2.04)	3.77 (0.46-31.14)	1.67 (0.42-6.60)	1.51 (0.33-7.00)
Nodule-in-nodule	HCC	1.87e+06 (0.00-Inf)	**2.32 (1.06-5.08)***	**5.88 (3.08-11.23)***	**2.39 (1.12-5.12)***	**2.57 (1.23-5.39)***	**3.35 (1.67-6.72)***	1.62 (0.75-3.48)
Mosaic architecture	HCC	2.48 (0.49-12.49)	2.17 (0.98-4.81)	**5.70 (2.97-10.94)***	**2.51 (1.17-5.38)***	**2.51 (1.19-5.30)***	**3.38 (1.68-6.79)***	1.55 (0.71-3.36)
Blood products in mass	HCC	1.13 (0.36-3.53)	**2.34 (1.06-5.15)***	**5.96 (3.10-11.46)***	**2.43 (1.12-5.27)***	**2.60 (1.23-5.49)***	**3.27 (1.62-6.59)***	1.60 (0.74-3.45)
Fat in mass, more than adjacent liver	HCC	1.05 (0.42-2.63)	**2.36 (1.08-5.17)***	**5.94 (3.08-11.47)***	**2.41 (1.12-5.15)***	**2.59 (1.23-5.46)***	**3.25 (1.61-6.54)***	1.60 (0.74-3.44)
Size stability > 2 years	Benignity	0.35 (0.10-1.24)	**3.38 (1.10-10.38)***	**6.33 (2.48-16.15)***	4.05 (0.79-20.73)	1.77 (0.35-9.00)	**3.39 (1.31-8.83)***	1.36 (0.49-3.79)
Size reduction	Benignity	0.24 (0.01-4.40)	**3.39 (1.23-9.30)***	**4.11 (1.82-9.28)***	**3.57 (1.08-11.76)***	2.32 (0.69-7.82)	**3.82 (1.66-8.78)***	1.29 (0.53-3.16)
Parallels blood pool	Benignity	**No positive observations**	**2.34 (1.07-5.12)***	**5.96 (3.12-11.40)***	**2.41 (1.13-5.16)***	**2.55 (1.22-5.36)***	**3.44 (1.72-6.91)***	1.60 (0.74-3.45)

Undistorted vessels	Benignity	1.20 (0.29-4.97)	**2.31 (1.05-5.08)***	**5.95 (3.12-11.38)***	**2.38 (1.11-5.12)***	**2.53 (1.20-5.33)***	**3.44 (1.71-6.90)***	1.62 (0.75-3.50)
Iron in mass, more than liver	Benignity	3.07e-07 (0.00-Inf)	2.14 (0.90-5.10)	**8.29 (3.81-18.06)***	2.14 (0.74-6.19)	1.76 (0.76-4.05)	2.12 (0.89-5.02)	1.18 (0.44-3.18)
Marked T2 hyperintensity	Benignity	0.49 (0.15-1.62)	**2.46 (1.12-5.42)***	**6.03 (3.13-11.61)***	**2.49 (1.16-5.35)***	**2.83 (1.32-6.09)***	**3.21 (1.59-6.47)***	1.55 (0.72-3.34)
Hepatobiliary phase iso-intensity	Benignity	---	---	---	---	---	---	---

*Significant association ($p<0.05$); Abbreviations- APHE: arterial phase hyperenhancement; HCC: hepatocellular carcinoma; US: ultrasound; CI: confidence interval

Appendix 3.11. Multivariable analysis of each ancillary feature, adjusted for all major features (APHE, size, nonperipheral washout, enhancing capsule and threshold growth), to determine the odds of malignancy.

Ancillary Feature	Association with (Based on LIRADS v2018)	OR (95% CI)	Enhancing capsule (95% CI)	APHE (95% CI)	Size 10-19 mm (95% CI)	Size ≥ 20mm (95% CI)	Nonperipheral washout (95% CI)	Threshold Growth (95% CI)
US visibility as discrete nodule	Malignancy	0.95 (0.13-7.10)	1.51e+08 (0.00-Inf)	0.88 (0.12-6.20)	3.89e-09 (0.00-Inf)	1.01e-08 (0.00-Inf)	2.26 (0.23-22.59)	0.36 (0.03-3.70)
Subthreshold growth	Malignancy	**11.92 (1.43-99.60)***	3.31 (0.80-13.65)	1.79 (0.65-4.89)	2.52 (0.64-9.99)	3.78 (0.84-17.02)	**4.08 (1.32-12.59)***	1.01 (0.25-4.03)
Restricted diffusion	Malignancy	**4.03 (1.42-11.46)***	**4.29 (1.08-16.95)***	1.50 (0.62-3.61)	**0.25 (0.09-0.72)***	**0.21 (0.07-0.67)***	**4.73 (1.64-13.66)***	1.32 (0.41-4.24)
Mild-moderate T2 hyperintensity	Malignancy	**3.42 (1.44-8.13)***	**4.77 (1.15-19.81)***	1.30 (0.53-3.18)	0.49 (0.18-1.37)	0.36 (0.12-1.08)	**4.50 (1.62-12.46)***	2.07 (0.65-6.63)
Corona enhancement	Malignancy	8.52e+06 (0.00-Inf)	**4.78 (1.23-18.50)***	1.08 (0.47-2.47)	0.45 (0.17-1.17)	0.45 (0.16-1.25)	**3.93 (1.53-10.12)***	2.11 (0.70-6.37)
Fat sparing in solid mass	Malignancy	0.47 (0.08-2.85)	**4.48 (1.02-19.68)***	1.23 (0.44-3.44)	0.31 (0.09-1.12)	**0.18 (0.05-0.67)***	**3.99 (1.14-14.02)***	1.54 (0.39-6.10)
Iron sparing in solid mass	Malignancy	1.86 (0.34-10.28)	3.92 (0.97-15.83)	1.73 (0.60-4.97)	**0.04 (0.004-0.33)***	**0.03 (0.003-0.24)***	2.83 (0.79-10.19)	1.41 (0.35-5.74)
Transitional phase hypointensity	Malignancy	---	---	---	---	---	---	---
Hepatobiliary phase hypointensity	Malignancy	---	---	---	---	---	---	---
Nonenhancing "capsule"	HCC	1.57e-32 (0.00-Inf)	0.12 (0.00-Inf)	3.78e+30 (0.00-Inf)	7.60e-04 (0.00-Inf)	1.28 (0.00-Inf)	0.34 (0.00-Inf)	0.63 (0.00-Inf)
Nodule-in-nodule	HCC	3.35e+06 (0.00-Inf)	**4.81 (1.23-18.76)***	1.06 (0.46-2.44)	0.43 (0.16-1.12)	0.43 (0.15-1.19)	**3.73 (1.43-9.72)***	2.08 (0.69-6.29)
Mosaic architecture	HCC	1.90e+07 (0.00-Inf)	**4.40 (1.12-17.34)***	1.00 (0.44-2.30)	0.45 (0.17-1.18)	0.40 (0.14-1.13)	**3.84 (1.48-9.98)***	1.96 (0.64-5.99)
Blood products in mass	HCC	0.83 (0.17-4.05)	**5.04 (1.27-19.99)***	1.15 (0.50-2.66)	0.42 (0.16-1.12)	0.38 (0.14-1.09)	**3.98 (1.52-10.39)***	2.12 (0.69-6.48)
Fat in mass, more than adjacent liver	HCC	0.47 (0.16-1.41)	**5.16 (1.31-20.36)***	1.25 (0.53-2.93)	0.39 (0.15-1.05)	0.37 (0.13-1.07)	**4.33 (1.63-11.53)***	2.20 (0.71-6.81)
Size stability > 2 years	Benignity	0.45 (0.12-1.67)	**4.21 (1.04-17.08)***	**3.60 (1.33-9.78)***	3.99 (0.80-19.90)	3.67 (0.71-18.98)	**3.58 (1.25-10.28)***	1.29 (0.42-3.99)
Size reduction	Benignity	0.15 (0.01-3.96)	**4.59 (1.19-17.78)***	**2.65 (1.08-6.53)***	**3.64 (1.09-12.18)***	**4.34 (1.21-15.51)***	**3.84 (1.48-9.96)***	1.53 (0.54-4.32)
Parallels blood pool	Benignity	**No positive observations**	**4.77 (1.23-18.58)***	1.08 (0.47-2.48)	0.43 (0.17-1.13)	0.42 (0.15-1.18)	**3.90 (1.51-10.11)***	2.05 (0.68-6.22)
Undistorted vessels	Benignity	1.31 (0.20-8.75)	**4.69 (1.20-18.34)***	1.08 (0.47-2.48)	0.42 (0.16-1.12)	0.42 (0.15-1.17)	**3.88 (1.50-10.07)***	2.09 (0.68-6.36)

Iron in mass, more than liver	Benignity	7.39e-09 (0.00-Inf)	5.18 (0.99-26.99)	1.33 (0.43-4.08)	**0.04 (0.005-0.37)***	**0.04 (0.004-0.32)***	2.63 (0.69-10.07)	1.18 (0.22-6.31)
Marked T2 hyperintensity	Benignity	0.34 (0.08-1.41)	**5.46 (1.37-21.76)***	1.13 (0.49-2.61)	0.47 (0.18-1.23)	0.45 (0.16-1.31)	**3.84 (1.47-10.00)***	2.01 (0.66-6.15)
Hepatobiliary phase iso-intensity	Benignity	---	---	---	---	---	---	---

*Significant association ($p<0.05$); Abbreviations- APHE: arterial phase hyperenhancement; HCC: hepatocellular carcinoma; US: ultrasound; CI: confidence interval

Appendix 3.12. The within study (I^2) and between study-variance (tau^2) in each random effects model, after adjusting for all major features.

Ancillary Feature	HCC		Malignancy	
	I^2 [%(95%CI)]	**Tau^2**	**I^2 [%(95%CI)]**	**Tau^2**
US visibility as discrete nodule	0.0	0	0.0	0
Subthreshold growth	0.0 (0.0-89.6)	0	0.0 (0.0-89.6)	0
Restricted diffusion	86.4 (66.8-94.4)	0.302	0.0 (0.0-84.7)	2.344
Mild-moderate T2 hyperintensity	86.4 (67.0-94.4)	0.287	0.0 (0.0-84.7)	2.376
Corona enhancement	86.3 (66.5-94.4)	0.282	0.0 (0.0-84.7)	2.331
Fat sparing in solid mass	80.8 (49.6-92.7)	0.228	21.1 (0.0-87.9)	2.831
Iron sparing in solid mass	82.7 (55.7-93.3)	0.349	0.0 (0.0-84.7)	1.883
Transitional phase hypointensity	---	---	---	---
Hepatobiliary phase hypointensity	---	---	---	---
Nonenhancing "capsule"	85.5 (41.5-96.4)	0.468	0.0	4.251
Nodule-in-nodule	86.3 (66.5-94.4)	0.282	0.0 (0.0-84.7)	2.331
Mosaic architecture	86.3 (66.5-94.4)	0.282	0.0 (0.0-84.7)	2.331
Blood products in mass	86.4 (67.0-94.4)	0.287	0.0 (0.0-84.7)	2.376
Fat in mass, more than adjacent liver	86.4 (67.0-94.4)	0.287	0.0 (0.0-84.7)	2.376
Size stability > 2 years	29.4 (0.0-92.7)	0.015	0.0 (0.0-89.6)	0
Size reduction	0.0 (0.0-89.6)	0	0.0 (0.0-89.6)	0
Parallels blood pool	86.3 (66.5-94.4)	0.282	0.0 (0.0-84.7)	2.331
Undistorted vessels	86.3 (66.5-94.4)	0.282	0.0 (0.0-84.7)	2.331
Iron in mass, more than liver	86.5 (61.2-95.3)	0.322	0.0 (0.0-89.6)	3.27
Marked T2 hyperintensity	86.4 (67.0-94.4)	0.287	0.0 (0.0-84.7)	2.376
Hepatobiliary phase iso-intensity	---	---	---	---

Abbreviations- US: ultrasound