Title

Diagnostic Performance and Interreader Agreement of a Standardized MR Imaging Approach in the Prediction of Small Renal Mass Histology.

Journal

Radiology, 287(2)

ISSN

0033-8419

Authors

Kay, Fernando U
Canvasser, Noah E
Xi, Yin
et al.

Publication Date

2018-05-01

DOI

10.1148/radiol.2018171557

Peer reviewed

Diagnostic Performance and Interreader Agreement of a Standardized MR Imaging Approach in the Prediction of Small Renal Mass Histology[1]

Fernando U. Kay, MD
Noah E. Canvasser, MD
Yin Xi, PhD
Daniella F. Pinho, MD
Daniel N. Costa, MD
Alberto Diaz de Leon, MD
Gaurav Khatri, MD
John R. Leyendecker, MD
Takeshi Yokoo, MD, PhD
Aaron H. Lay, MD
Nicholas Kavoussi, MD
Ersin Koseoglu, MD
Jeffrey A. Cadeddu, MD
Ivan Pedrosa, MD, PhD

Purpose: To assess the diagnostic performance and interreader agreement of a standardized diagnostic algorithm in determining the histologic type of small (≤4 cm) renal masses (SRMs) with multiparametric magnetic resonance (MR) imaging.

Materials and Methods: This single-center retrospective HIPAA-compliant institutional review board–approved study included 103 patients with 109 SRMs resected between December 2011 and July 2015. The requirement for informed consent was waived. Presurgical renal MR images were reviewed by seven radiologists with diverse experience. Eleven MR imaging features were assessed, and a standardized diagnostic algorithm was used to determine the most likely histologic diagnosis, which was compared with histopathology results after surgery. Interreader variability was tested with the Cohen κ statistic. Regression models using MR imaging features were used to predict the histopathologic diagnosis with 5% significance level.

Results: Clear cell renal cell carcinoma (RCC) and papillary RCC were diagnosed, with sensitivities of 85% (47 of 55) and 80% (20 of 25), respectively, and specificities of 76% (41 of 54) and 94% (79 of 84), respectively. Interreader agreement was moderate to substantial (clear cell RCC, κ = 0.58; papillary RCC, κ = 0.73). Signal intensity (SI) of the lesion on T2-weighted MR images and degree of contrast enhancement (CE) during the corticomedullary phase were independent predictors of clear cell RCC (SI odds ratio [OR]: 3.19; 95% confidence interval [CI]: 1.4, 7.1; $P = .003$; CE OR, 4.45; 95% CI: 1.8, 10.8; $P < .001$) and papillary RCC (CE OR, 0.053; 95% CI: 0.02, 0.2; $P < .001$), and both had substantial interreader agreement (SI, κ = 0.69; CE, κ = 0.71). Poorer performance was observed for chromophobe histology, oncocytomas, and minimal fat angiomyolipomas, (sensitivity range, 14%–67%; specificity range, 97%–99%), with fair to moderate interreader agreement (κ range = 0.23–0.43). Segmental enhancement inversion was an independent predictor of oncocytomas (OR, 16.21; 95% CI: 1.0, 275.4; $P = .049$), with moderate interreader agreement (κ = 0.49).

Conclusion: The proposed standardized MR imaging–based diagnostic algorithm had diagnostic accuracy of 81% (88 of 109) and 91% (99 of 109) in the diagnosis of clear cell RCC and papillary RCC, respectively, while achieving moderate to substantial interreader agreement among seven radiologists.

Online supplemental material is available for this article.

[1] From the Department of Radiology (F.U.K., Y.X., D.F.P., D.N.C., A.D.d.L., G.K., J.R.L., T.Y., J.A.C., I.P.), Department of Urology (N.E.C., A.H.L., N.K., E.K., J.A.C., I.P.), and Advanced Imaging Research Center (D.C., T.Y., I.P.), University of Texas Southwestern Medical Center, 2201 Inwood Rd, Suite NE2.210, Dallas, TX 75390-9085. Received July 10, 2017; revision requested September 7; revision received September 20; accepted September 28; final version accepted November 14. **Address correspondence to** I.P. (e-mail: *ivan.pedrosa@utsouthwestern.edu*).

Supported by the National Institutes of Health (5R01CA154475, P50CA196516).

The incidence of small renal masses (SRMs) (ie, T1a ≤4 cm) has dramatically increased in the past few decades, with the increase being mainly due to the widespread use of cross-sectional imaging (1). A shift toward earlier diagnosis of incidental SRMs has not translated, however, into lower renal cancer–specific mortality rates (2), suggesting a possible overtreatment effect. Indeed, up to 16% of SRMs are benign (3) and will not benefit from definitive therapy (ie, surgery or percutaneous ablation).

Management decisions range from definitive therapy to active surveillance (ie, systematic observation) and are guided by the perceived risk of disease progression and development of metastatic disease, which in turn, could be assessed via percutaneous biopsy with histopathologic analysis (4,5). Although percutaneous biopsy has been proposed for use in risk stratification (6), it is invasive and subject to complications (7). In addition, renal biopsy may yield indeterminate results (8), particularly for tumor grading (9). Thus, the development of competitive noninvasive methods with which to assess histologic type and tumor grade would be desirable.

Numerous studies have reported the potential of magnetic resonance (MR) imaging in the histologic subtyping of renal cell carcinoma (RCC) and in the differentiation of benign from malignant renal masses. For instance, T2-weighted (T2W) MR imaging aids in the differentiation of *(a)* papillary RCC from clear cell RCC (10) and *(b)* minimal fat angiomyolipoma (AML) from clear cell RCC (11). Similarly, multiphasic and dynamic contrast material–enhanced MR imaging can help differentiate between clear cell, papillary, and chromophobe RCC (12) and between oncocytomas and chromophobe or clear cell RCC (13). Despite the existence of robust data supporting the utility of MR imaging features in the characterization of SRM, there is a paucity of initiatives aimed at creating an easily applicable, standardized, and robust diagnostic system for multiparametric MR imaging in clinical practice. The purpose of our study was to determine the diagnostic performance and interreader agreement of a standardized diagnostic algorithm used to determine the histologic type of SRM at multiparametric MR imaging.

Implication for Patient Care

- The implementation of a standardized diagnostic algorithm has the potential to yield superior interreader agreement and diagnostic accuracy for histologic subtyping of small renal masses (ie, T1a disease) with multiparametric MR imaging.

Materials and Methods

Study Design

This single-center retrospective Health Insurance Portability and Accountability Act–compliant study was approved by the institutional review board, and the requirement for informed consent was waived. We reviewed medical records to identify patients who underwent surgical resection of a renal mass measuring 4 cm or less between December 2011 and July 2015 and who had a presurgical multiparametric MR imaging study of the kidneys available for analysis. Patients with a pathologic stage of T1b or greater, those with uninterpretable MR images due to poor image quality, and those in whom essential MR imaging sequences were not performed per local institutional protocol (described later) were excluded (Fig 1). This cohort was included in a previous study, although no overlap exists in the analyses performed (14). The previous study evaluated the diagnostic accuracy of a likelihood score for the prediction of clear cell histology (14). In contrast, the current study assesses the diagnostic performance and interreader agreement of a standardized diagnostic algorithm for the diagnosis of the most common malignant and benign histologic diagnoses in SRM and analyzes the contribution of individual imaging findings in achieving the correct diagnosis.

MR Imaging

Essential sequences included axial and coronal non–fat-suppressed and fat-suppressed T2W multishot or single-shot fast spin-echo imaging, T1-weighted (T1W) gradient-recalled-echo in- and opposed-phase imaging, and three-dimensional fat-suppressed spoiled gradient-recalled-echo or Dixon-based multiphasic contrast-enhanced T1W imaging. Acquisition parameters are available in Table E1 (online). Studies obtained at different medical centers but with the essential sequences were included. One of the body MR imaging–trained authors who did not analyze imaging features (F.K., 9 years of postresidency practice) was responsible for reviewing all studies to exclude those with suboptimal imaging quality or nonstandard acquisition parameters. All patients who underwent imaging at our institution received a bolus of 0.1 mmol per kilogram of body weight of intravenous gadobutrol (Gadavist; Bayer Healthcare). Contrast-enhanced acquisitions were timed to the corticomedullary phase by using MR fluoroscopy. The early and late nephrographic phases and the early and late

https://doi.org/10.1148/radiol.2018171557

Content code: GU

Radiology 2018; 287:543–553

Abbreviations:
AML = angiomyolipoma
RCC = renal cell carcinoma
SRM = small renal mass
T1W = T1 weighted
T2W = T2 weighted

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

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

Figure 1

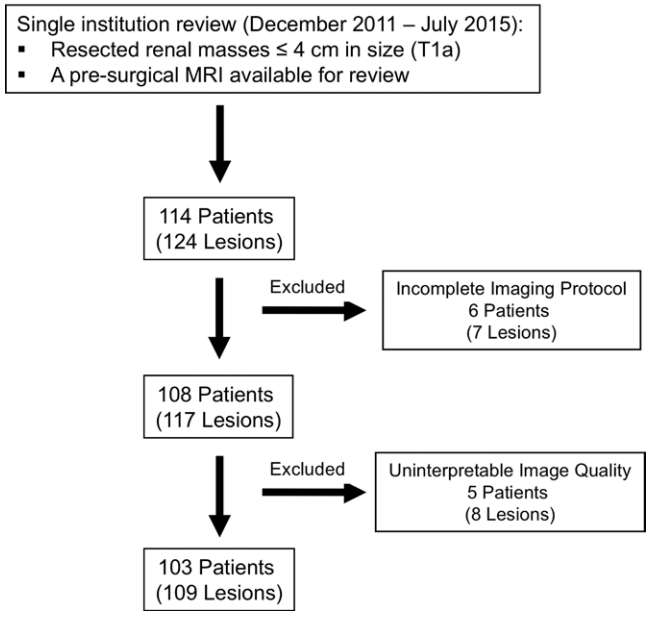

Figure 1: Study flowchart.

excretory phases were acquired 20 seconds, 40 seconds, 2 minutes, and 5 minutes after the corticomedulary phase, respectively. All MR imaging studies obtained at other imaging centers and included in this study had acquisitions performed in both the corticomedullary phase and the late nephrographic phase. Although diffusion-weighted images were also acquired as part of this protocol at the authors' institution, they were not consistently available in studies obtained elsewhere; thus, diffusion-weighted imaging was not included in the analysis.

Image Interpretation

Images were independently reviewed by seven experienced radiologists with body MR imaging training (I.P., 15 years; J.R.L., 15 years; D.N.C., 12 years; G.K., 10 years; T.Y., 7 years; D.F.P., 5 years; A.D.L., 1 year), who were blinded to clinical and histopathologic information, using a picture archiving and communication workstation (iSite Intellispace; Philips, Best, the Netherlands). Reviewers completed a training session prior to image review to standardize image interpretation. The training session consisted of a slide presentation with examples of the different criteria proposed in the standardized diagnostic algorithm (detailed later) using renal masses not included in the cohort analyzed in this study. This algorithm was created by an author (I.P.) based on clinical experience and had been used as a teaching tool for residents and fellows of the body MR imaging section for 2 years before the study. All reviewers were familiarized with the tool. Table 1 summarizes the definition of each MR imaging feature with its respective categories.

After reviewers assessed the MR imaging features, they were asked to provide the most likely histopathologic diagnosis. A previously reported diagnostic algorithm (17) (Fig 2) was provided as a guideline for defining the most likely diagnosis; notwithstanding, reviewers were allowed to overrule the algorithm ad libitum. In addition, a nonspecific diagnosis could be assigned if imaging features failed to fall into a specific diagnostic bin. This option was allowed since some of the diagnostic branches yielded more than one diagnostic possibility (eg, isointense signal on T2W images and high contrast enhancement on contrast-enhanced T1W images); hence, a specific diagnosis could not be established.

Reference Standard

The postsurgical histologic diagnosis served as the reference standard for all tumors and was determined by reviewing the medical records. The diagnosis was performed by genitourinary pathologists at our institution who classified and graded all tumors (ie, RCC) by using the 2004 World Health Organization and International Society of Urological Pathology classifications (19,20).

Statistical Analyses

Performance analyses included only the most common diagnoses (ie, clear cell RCC, papillary RCC, chromophobe RCC, oncocytoma, minimal fat AML) because of the low frequency of other diagnoses; however, data from all lesions were tabulated. Uni- and multivariate analyses to predict histologic diagnosis were performed for each finding by using generalized linear mixed models with logit link function adjusted for lesions clustered in the same patient. To account for the multireader study design, reviewers were also treated as a random factor in all mixed-effect models. The dependent variable was dichotomized as the occurrence of the specific diagnosis versus all other diagnoses by using the postsurgical histologic reference standard. Variable selection at multivariable stepwise analysis was performed with L1-penalized estimation, which shrinks the coefficients of the variables per a penalty parameter (λ). The optimal penalty parameter was chosen based on the Akaike information criterion. Coefficients of the L1-penalized estimation regression were obtained, and hypothesis tests with the null of no effect (coefficient = 0) were conducted with a z test. Diagnostic performance was tested against the histopathologic diagnosis for each reviewer, with the median and ranges among reviewers reported for sensitivity, specificity, positive and negative predictive values, and accuracy.

Table 1

MR Imaging Features: Definition of Imaging Features and Respective Categories

Feature	Definition	Categories
T2W signal intensity	Lesion intensity in comparison with intensity of the normal renal cortex on non–fat-suppressed T2W images	1, low; 2, isointense; 3, high
T2W signal heterogeneity	Lesion texture as observed on fat-suppressed T2W images	1, homogeneous; 2, minimally heterogeneous; 3, heterogeneous
Intravoxel fat	Signal dropout on opposed-phase images compared with that on in-phase images	1, absent; 2, present
Bulk fat	High signal intensity on non–fat-suppressed T1W or T2W images, with signal dropout on fat-suppressed images; India ink artifact at the interface between the hyperintense focus and adjacent renal parenchyma	1, absent; 2, present
Magnetic susceptibility	Decreased signal intensity on T1W in-phase images compared with that on the shorter-echo-time opposed-phase images	1, absent; 2, present
Central scar	High signal intensity on T2W images at the lesion core, exhibiting no initial enhancement after intravenous contrast material administration and surrounded by enhancing solid tumor; it can enhance on delayed images	1, absent; 2, present
Hemorrhage	High signal intensity on unenhanced fat-saturated T1W images	1, absent; 2. present
Segmental enhancement inversion	Areas within the renal mass with intense enhancement during the corticomedullary phase, later washout, and areas with low-level enhancement during the corticomedullary phase exhibiting intense enhancement on delayed contrast-enhanced images (15,16)	1, absent; 2, present
Contrast enhancement	During the corticomedullary phase, which was rated as low, moderate, or high if the enhancing portions of the renal mass exhibited approximately ≤30%, 50%, or 100% enhancement, respectively, compared with renal cortex enhancement	1, low; 2, moderate; 3, high
Dynamic characteristics	Mass enhancement during the late nephrographic phase, as follows: progressive (at least 10% more than in the corticomedullary phase), plateau (approximately 10% of the corticomedullary phase), or washout (at least 10% less than in the corticomedullary phase)	1, progressive; 2, plateau; 3, washout
Enhancement heterogeneity	Lesion texture as observed on contrast-enhanced images	1, homogeneous; 2, minimally heterogeneous; 3, heterogeneous

Note.—T1W = T1 weighted, T2W = T2 weighted.

Overall differences in diagnostic accuracy when reviewers adhered to the proposed diagnostic algorithm versus nonadherence (ie, final diagnosis not included in the branch of the diagnostic algorithm with the selected specific MR imaging features) were tested with the Cochran-Mantel-Haenszel test. The simple pairwise Cohen κ statistic was used to measure interreader agreement for binary response imaging features, whereas the linear weighted pairwise κ statistic was used to assess agreement for ordinal ternary response features (κ = 0–0.20, slight agreement; κ = 0.21–0.40, fair agreement; κ = 0.41–0.60, moderate agreement; κ = 0.61–0.80, substantial agreement; and κ = 0.81–1, almost perfect agreement). Spearman ρ was used to evaluate the relationship between years of experience and diagnostic performance. All statistical analyses were performed with SAS, version 9.4 software (SAS Institute, Cary, NC). $P < .05$ indicated a significant difference.

Results

Study Cohort

The final cohort comprised 103 patients with 109 lesions (47 women [46%], 56 men [54%]; mean age, 56.7 years ± 14.1; mean body mass index, 28.7 kg/m^2 ± 5.9). Nine (9%) patients had a previous history of RCC different than the SRM included in this study. The median time between MR imaging and surgery was 65 days (range, 12–486 days). Approximately half of the renal masses were imaged at our institution (54 of 109), with the remainder of studies (n = 55) acquired at outside facilities. Imaging was performed with a 1.5-T system in 61 (59%) patients and with a 3.0-T system in 42 (41%). MR imager manufacturers included Philips (46%, n = 47), Siemens (Erlangen, Germany) (34%, n = 35), General Electric (Waukesha, Wis) (17%, n = 18), Hitachi (Tokya, Japan) (2%, n = 2), and Toshiba (Otawara, Japan) (1%, n = 1). Table 2 includes frequencies and sizes of all renal masses included in the study. Clear cell RCC was the most common diagnosis (50%, n = 55), followed by papillary RCC (23%, n = 25) and chromophobe RCC (6%, n = 7). Benign histology accounted for 16% (n = 17) of the lesions, with oncocytoma (n = 6) and minimal fat AML (n = 6) being the most common benign lesions.

Figure 2

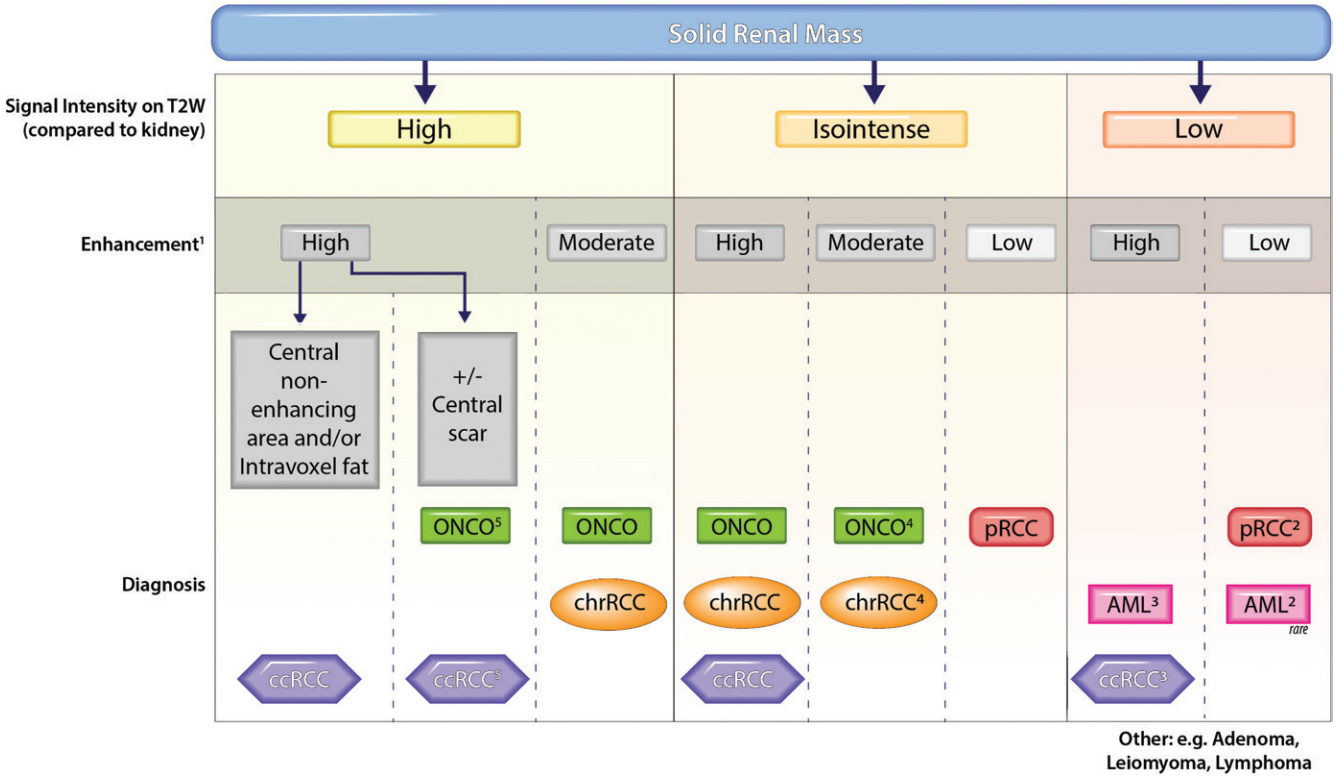

Figure 2: Diagnostic algorithm for characterization of solid renal masses. Vertical dotted lines indicate different branches in the diagnostic algorithm based on T2-weighted and contrast enhancement characteristics. The most likely diagnoses in each branch are indicated at the bottom of each column. Readers use ancillary findings to select the most likely diagnosis within each branch. *1* = Enhancement during the corticomedullary phase was categorized as high (enhancement was the same as or greater than that in the renal cortex), moderate (approximately 50% of that in the renal cortex), or low (approximately 25%–30% of that in the renal cortex). *2* = Arterial-to-delayed enhancement ratio is the difference in signal intensity between the arterial and precontrast phases divided by the difference in signal intensity between the delayed and precontrast phases (18). Arterial-to-delayed enhancement ratio greater than 1.5 favors a diagnosis of minimal fat AML, while an arterial-to-delayed enhancement ratio less than 1.5 favors a diagnosis of renal cell carcinoma (RCC). *3* = Clear cell RCC is typically heterogeneous, whereas minimal fat AML is typically homogeneous. *4* = Oncocytoma is more commonly hypervascular (enhances similarly to renal cortex), while chromophobe RCC more commonly has moderate enhancement (approximately 50% of renal cortex). *5* = A diagnosis of oncocytoma is made if central scar (ie, central initial nonenhancing area with delayed enhancement) is present, whereas diagnosis of clear cell RCC is favored if necrosis (ie, nonenhancing central area) is present or if the tumor is heterogeneous. AML = angiomyolipoma, ccRCC = clear cell RCC, chrRCC = chromophobe RCC, ONCO = oncocytoma, pRCC = papillary RCC.

Interreader Agreement, MR Imaging Features, and Histologic Diagnosis

Median pairwise interreader agreement was substantial for degree of contrast enhancement (κ = 0.71), T2W signal intensity (κ = 0.69), presence of hemorrhage (κ = 0.67), and bulk fat (κ = 0.66). Interreader agreement was moderate for the presence of magnetic susceptibility (κ = 0.57), segmental enhancement inversion (κ = 0.49), T2W texture (κ = 0.48), intravoxel fat (κ = 0.47), and central scar (κ = 0.42). Agreement was only fair for enhancement homogeneity (κ = 0.35) and dynamic characteristics (κ = 0.28).

Diagnostic performance (sensitivity, specificity, positive and negative predictive values, and accuracy) for reviewers' most likely histopathologic diagnosis at multiparametric MR imaging and respective interreader agreements is summarized in Table 3. Median pairwise interreader agreement with the algorithm was substantial for papillary RCC (κ = 0.73), moderate for clear cell RCC (κ = 0.58) and minimal fat AML (κ = 0.43), and fair for oncocytoma (κ = 0.26) and chromophobe RCC (κ = 0.23).

Table 4 shows the results from uni- and multivariate models using MR imaging features to predict clear cell RCC, papillary RCC, chromophobe RCC, oncocytoma, and minimal fat AML.

Clear Cell RCC

Features associated with clear cell RCC at univariate analyses included increased T2W signal intensity (odds ratio [OR], 3.29; 95% confidence interval [CI]: 2.1, 5.2; $P < .001$), heterogeneous T2W texture (OR, 3.52; 95% CI: 1.9, 6.4; $P < .001$), intravoxel fat (OR, 2.68; 95% CI: 1.3, 5.8; $P = .011$), high

contrast enhancement (OR, 9.99; 95% CI: 4.6, 21.7; $P < .001$), and heterogeneous enhancement (OR, 3.56; 95% CI: 2.1, 6.2; $P < .001$). At multivariate analysis, clear cell RCC was predicted by signal intensity on T2W images (high vs low: OR, 3.19; 95% CI: 1.4, 7.1; $P < .001$) and degree of contrast enhancement (high vs low: OR, 4.45; 95% CI: 1.8, 10.8; $P < .001$). Sensitivities and specificities of reviewers for the diagnosis of clear cell RCC ranged from 71% (39 of 55) to 98% (54 of 55) and from 69% (37 of 54) to 85% (46 of 54), respectively. Fig 3a–3d exemplifies MR imaging features encountered in a pathologically proven clear cell RCC in our data set.

Papillary RCC

Features associated with papillary RCC were lower T2W signal intensity (OR, 0.19; 95% CI: 0.06, 0.6; $P = .006$), homogeneous T2W texture (OR, 0.34; 95% CI: 0.1, 0.8; $P = .014$), absence of central scar (OR, 0.11; 95% CI: 0.01, 0.8; $P = .033$), presence of hemorrhage (OR, 5.45; 95% CI: 1.3, 23.6; $P = .024$), low contrast enhancement (OR, 0.056; 95% CI: 0.01, 0.3; $P < .001$), and homogeneous contrast enhancement (OR, 0.29; 95% CI: 0.09, 0.9; $P = .03$) at univariate analyses. Papillary RCC was only independently predicted by degree of contrast enhancement (low vs high: OR, 0.053; 95% CI: 0.02, 0.2; $P < .001$) at multivariate analysis (Fig 4a, 4b). The most likely histopathologic diagnosis defined by the reviewers had sensitivities that ranged from 68% (17 of 25) to 96% (24 of 25) and specificities that ranged from 90% (76 of 84) to 98% (82 of 84) in the diagnosis of papillary RCC.

Chromophobe RCC

Features associated with chromophobe RCC at univariate analyses included low T2W signal intensity (OR, 0.56; 95% CI: 0.31, 0.99; $P = .046$) and absence of intravoxel fat (OR, 0.045; 95% CI: 0.006, 0.3; $P = .003$). None of the MR imaging features could be used to independently predict chromophobe RCC at multivariate analysis. Reviewer's sensitivity in the diagnosis of chromophobe RCC ranged from 14% (one of seven) to 29% (two of seven), while specificity ranged from 95% (97 of 102) to 100% (102 of 102).

Oncocytoma

Features associated with oncocytoma were absence of intravoxel fat (OR, 0.18; 95% CI: 0.03, 0.98; $P = .047$), presence of central scar (OR, 6.28; 95% CI: 2.2, 17.7; $P = .001$), presence of segmental enhancement inversion (OR, 28.50; 95% CI: 5.1, 160.9; $P < .001$), and higher contrast enhancement (OR, 3.58; 95% CI: 1.7, 7.5; $P = .001$) at univariate analyses. Segmental

Table 2

Tumor Size and Histologic Type

Pathologic Diagnosis	No. of Tumors (%)	Tumor Diameter (mm)*
Clear cell RCC	55 (50.5)	22.4 ± 8.18
Low-grade clear cell RCC	49 (45)	21.51 ± 6.45
High-grade clear cell RCC	6 (5.5)	29.67 ± 9.22
Papillary RCC	25 (22.9)	22.94 ± 7.33
Type 1	20 (18.3)	25.15 ± 8.69
Type 2	5 (4.6)	25.00 ± 4.30
Chromophobe RCC	7 (6.4)	23.71 ± 8.87
Oncocytoma	6 (5.5)	19.83 ± 4.26
Minimal fat angiomyolipoma	6 (5.5)	19.33 ± 8.16
Other	10 (9.2)	21.60 ± 8.04
Clear cell papillary RCC	3 (2.8)	22.95 ± 7.30
Multilocular benign cyst	2 (1.8)	21.67 ± 8.31
Cystic nephroma	2 (1.8)	24.60 ± 6.84
Metanephric adenoma	1 (0.9)	28
Tuberous sclerosis RCC	1 (0.9)	22
Carcinoid	1 (0.9)	14
Total	109 (100)	22.72 ± 7.42

Note.—Unless otherwise indicated, data are number of tumors and adata in parentheses are percentages. RCC = renal cell carcinoma.

* Data are mean ± standard deviation.

Table 3

Diagnostic Performance by Pathologic Subtype

Pathologic Subtype	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)	κ Value
Clear cell RCC	85 (71–98)	76 (69–85)	77 (74–87)	83 (74–98)	81 (75–92)	0.58 (0.33–0.71)
Papillary RCC	80 (68–96)	94 (90–98)	79 (71–92)	94 (91–98)	91 (88–97)	0.73 (0.61–0.81)
Chromophobe RCC	14 (14–29)	99 (95–100)	50 (17–100)	94 (94–95)	94 (90–95)	0.23 (-0.02–0.80)
Oncocytoma	33 (17–83)	97 (91–99)	38 (25–63)	96 (95–99)	94 (90–96)	0.26 (0.07–0.80)
Minimal fat AML	67 (33–83)	98 (95–100)	67 (44–100)	98 (95–98)	98 (95–98)	0.43 (0.19–0.65)

Note.—Data are medians among all reviewers, with the range in parentheses. AML = angiomyolipoma, NPV = negative predictive value, PPV = positive predictive value, RCC = renal cell carcinoma.

Table 4

Logistic Regression: Imaging Features by Histologic Subtype (All Reviewers)

	Clear Cell RCC		Papillary RCC		Chromophobe RCC		Oncocytoma		Minimal Fat AML	
Feature	Univariate	Multivariate	Univariate	Multivariate	Univariate	Multivariate	Univariate	Multivariate	Univariate	Multivariate
T2W signal intensity	OR, 3.29; $P < .001$	OR, 3.19; $P = .003$	OR, 0.19; $P = .006$	NS	OR, 0.56; $P = .046$	NS	NS	NS	OR, 0.37; $P = .001$	NS
T2W signal heterogeneity	OR, 3.52; $P < .001$	NS	OR, 0.34; $P = .014$	NS	NS	NS	NS	NS	OR, 0.39; $P = .003$	NS
Intravoxel fat	OR, 2.68; $P = .011$	NS	NS	NS	OR, 0.045; $P = .003$	NS	OR, 0.18; $P = .047$	NS	OR, 3.20; $P = .018$	NS
Bulk fat	NS	NS	NS	NS	NC	NS	NC	NS	OR, 80.14 $P = .047$	NS
Magnetic susceptibility	NS	NS	NS	NS	NS	NS	NC	NS	OR, 0.11; $P = .040$	NS
Central scar	OR, 2.92; $P = .020$	NS	OR, 0.11; $P = .033$	NS	NS	NS	OR, 6.28; $P = .001$	NS	NS	NS
Hemorrhage	NS	NS	OR, 5.45; $P = .024$	NS	NS	NS	NS	NS	OR, 0.066; $P = .002$	NS
Segmental enhancement inversion	NS	NS	NC	NS	NS	NS	OR, 28.50; $P < .001$	OR, 16.21; $P = .049$	NS	NS
Contrast enhancement	OR, 9.99; $P < .001$	OR, 4.45; $P < .001$	OR, 0.056; $P < .001$	OR, 0.053; $P < .001$	NS	NS	OR, 3.58; $P = .001$	NS	NS	NS
Dynamic characteristics	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Enhancement heterogeneity	OR, 3.56; $P < .001$	NS	OR, 0.29; $P = .030$	NS	NS	NS	NS	NS	OR, 0.44; $P = .008$	NS

Note.—Categories within MR imaging features are ordinally arranged in Table 1. AML = angiomyolipoma, NC = not calculated due to collinearity, NS = nonsignificant ($P > .05$), OR = odds ratio, RCC = renal cell carcinoma, T2W = T2 weighted.

enhancement inversion was the one independent predictor of oncocytoma (present vs absent, OR: 16.21; 95% CI: 1.0, 275.4; $P < .05$) at multivariate analysis (Fig E1a, E1b [online]). Sensitivities for oncocytoma ranged from 17% (one of six) to 83% (five of six) among reviewers, while median specificities ranged from 91% (94 of 103) to 99% (102 of 103).

Minimal Fat AML

Features associated with minimal fat AML at univariate analyses were low T2W signal intensity (OR, 0.37; 95% CI: 0.2, 0.7; $P = .001$), homogeneous T2W texture (OR, 0.39; 95% CI: 0.2, 0.7; $P = .003$), presence of intravoxel fat (OR, 3.20; 95% CI: 1.2, 8.4; $P = .018$) or bulk fat (OR, 80.14; 95% CI: 1.1, 6075.0; $P = .047$), absence of magnetic susceptibility (OR, 0.11; 95% CI: 0.01, 0.9; $P = .04$), absence of hemorrhage (OR, 0.066; 95% CI: 0.01, 0.4; $P = .002$), and homogeneous enhancement (OR, 0.44; 95% CI: 0.2, 0.8; $P = .008$). None of the MR imaging features could be used to independently predict the diagnosis of minimal fat AML at multivariate analysis. Sensitivities for minimal fat AML ranged from 33% (two of six) to 83% (five of six) among reviewers, while median specificities ranged from 95% (98 of 103) to 100% (103 of 103).

Effects of Adherence to the Diagnostic Algorithm

The average accuracy of the most likely histopathologic diagnosis category provided by reviewers when the assigned MR imaging features matched the proposed diagnostic algorithm in Figure 2 was 65% (416 of 636). In contrast, the average accuracy decreased to 43% (54 of 127) ($P < .001$) when the diagnoses chosen by reviewers did not match the output results pertaining to the respective set of input MR imaging features (ie, features in a specific branch in the diagnostic algorithm).

Nonspecific Diagnosis and Effect of Experience on Performance

The number of nonspecific diagnoses among reviewers ranged from 2.75% to 20.18%, with a median value of 6.42%. There was no correlation between reviewer's years of experience and sensitivity ($\rho = -0.17$, $P = .34$) or specificity ($\rho = 0.054$, $P = .76$) for all diagnoses. Similarly, there was no correlation between the number of nonspecific diagnoses and experience ($\rho = 0.16$, $P = .73$).

Discussion

Our study enables us to confirm consistent associations of some MR imaging features with specific histologic subtypes in SRMs. Like previous researchers (10,21,22), we found greater prevalence of high signal intensity on T2W images in patients with clear cell RCC, contrasting with lower signal intensity found in those with papillary

Figure 3

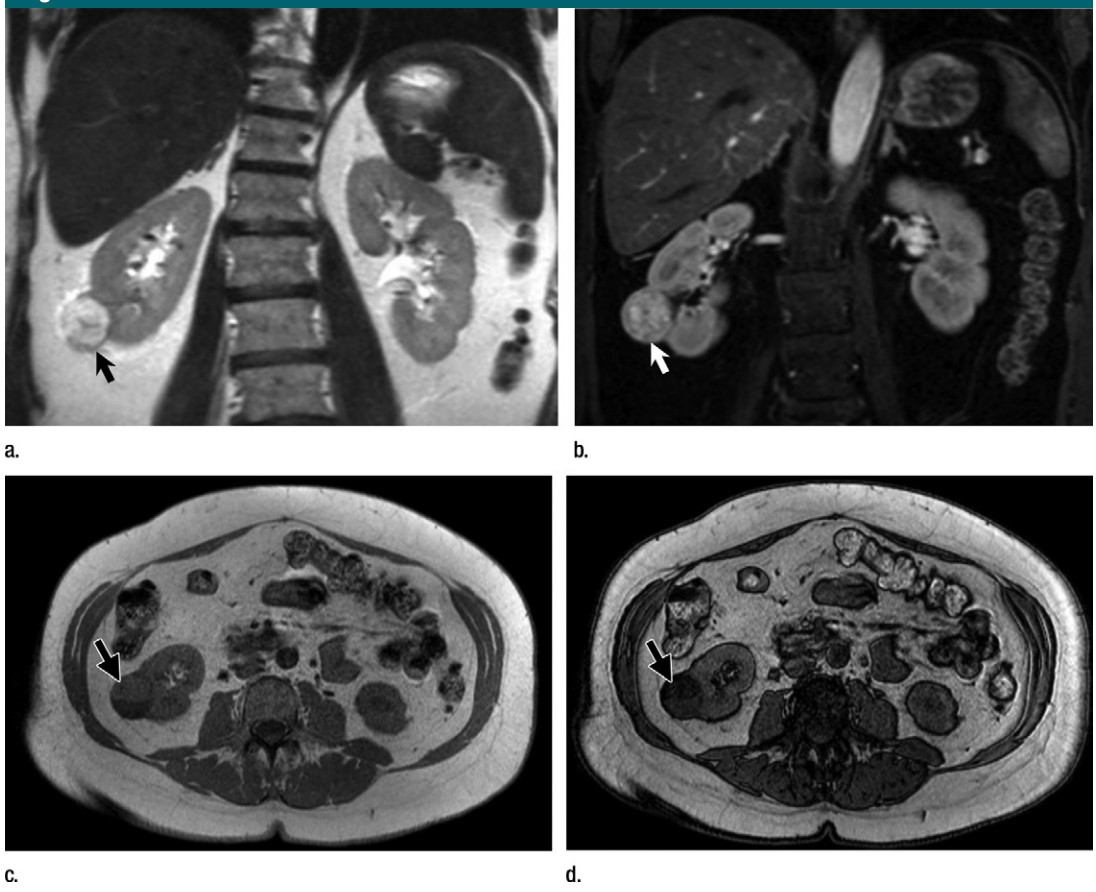

Figure 3: MR images show clear cell renal cell carcinoma (RCC). **(a)** Coronal non–fat-saturated T2-weighted single-shot fast spin-echo MR image shows a high-signal-intensity exophytic small renal mass (arrow) in the lower pole of the right kidney. **(b)** Coronal contrast-enhanced (corticomedullary) fat-saturated T1-weighted gradient-recalled-echo MR image shows the lesion (arrow) has high contrast enhancement. Axial **(c)** in-phase and **(d)** out-of-phase non–fat-saturated T1-weighted gradient-recalled-echo MR images. Note the presence of intravoxel fat, characterized by signal dropout, on **d** in comparison with **c** (arrows). Surgical resection enabled confirmation of clear cell RCC.

RCC, chromophobe RCC, or minimal fat AML. Similarly, higher contrast enhancement was most prevalent among patients with clear cell RCC and oncocytomas, as opposed to those with papillary RCC, who consistently showed lower contrast enhancement. Previously, Sun et al (12) found that contrast enhancement during the corticomedullary phase was a highly sensitive (93%) and specific (96%) MR imaging feature with which to differentiate clear cell RCC from papillary RCC with a quantitative method. Moreover, in our study, signal intensity on T2W images and degree of contrast enhancement on T1W images showed the highest interreader agreement among all MR imaging features.

Segmental enhancement inversion independently enabled prediction of oncocytomas in our cohort. Initial reports described segmental enhancement inversion as a reliable sign on computed tomographic (CT) images to distinguish oncocytomas from RCCs, with 80% sensitivity and 99% specificity (15). However, its validity has been disputed. Rosenkrantz et al (16) found segmental enhancement inversion in 29% of the oncocytomas, which was not statistically different from the 13% prevalence in chromophobe RCCs found in their cohort. The moderate interreader agreement for segmental enhancement inversion ($\kappa = 0.49$) in our study may explain to some extent the variability in these results.

Our proposed diagnostic algorithm was derived from published data and from the authors' experience, confirming T2W signal intensity and contrast enhancement as the main features with which to characterize SRM. The value of ancillary features, such as presence of intravoxel fat in clear cell RCC and minimal fat AML, signal heterogeneity in clear cell RCC, signal homogeneity in papillary RCC and minimal fat AML, and presence of hemorrhage in papillary RCC, were also analyzed. We found

Figure 4

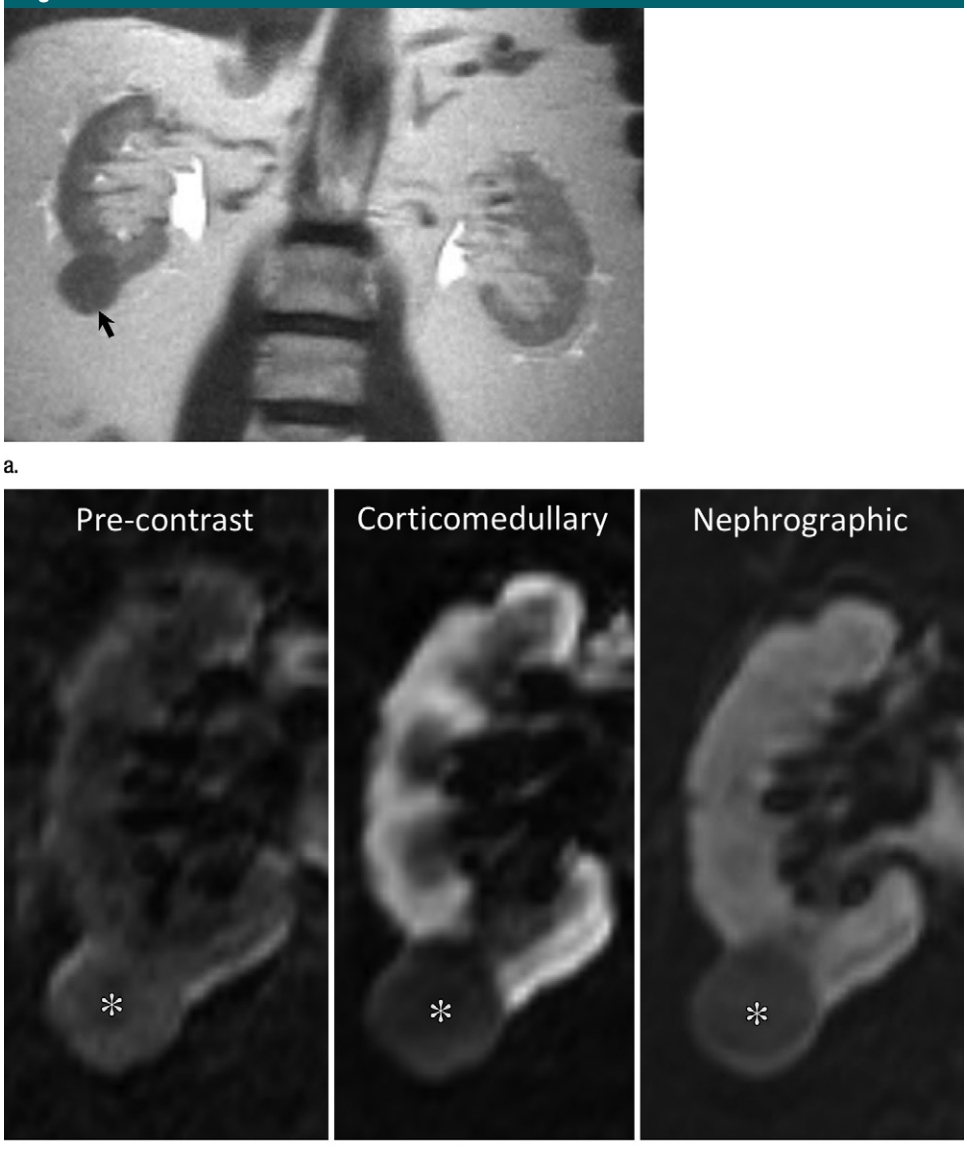

Figure 4: Papillary renal cell carcinoma (RCC). **(a)** Coronal non–fat-saturated T2-weighted single-shot fast spin-echo MR image shows a low-signal-intensity lesion (arrow). **(b)** Coronal T1-weighted fat-saturated spoiled-gradient recalled echo MR images show this same lesion with signal intensity analogous to that of the renal cortex, with low contrast enhancement on both corticomedullary and nephrographic phases (*) when compared with renal cortex enhancement. Pathologic analysis revealed papillary-type RCC.

a median sensitivity and specificity of 85% and 76%, respectively, for clear cell RCC, with a median sensitivity and specificity of 80% and 94%, respectively, for papillary RCC. The performance of this standardized approach is comparable with the results of Pedrosa et al (21), who used an MR imaging classification system for subtyping malignant renal masses, with a slightly higher sensitivity and specificity for clear cell RCC (92% and 83%, respectively), and similar sensitivity and specificity for papillary RCC (80% and 94%, respectively). Importantly, our cohort of lesions includes a mix of both benign and malignant renal masses representative of those routinely encountered in clinical practice, whereas most of the previous studies included only selected tumor types. Moreover, our data set comprised a defined group of tumors (SRM ≤4 cm, T1a), making our results more applicable to similar SRM elsewhere.

To our knowledge, our study assessing interreader agreement and diagnostic performance of MR imaging in the histologic subtyping of SRM included the most readers (with different levels of clinical expertise) to date. Interreader agreement in the diagnosis of clear cell RCC and papillary RCC varied from moderate to substantial in our cohort. By comparison, the Breast Imaging and Reporting Data System (23) and the Prostate Imaging and Reporting Data System (24), two validated diagnostic systems developed for use in mammographic screening and assessment of multiparametric prostate MR imaging, respectively, also show moderate to substantial agreement among reviewers. Cohen κ values of 0.51 to 0.59 are reported for Prostate Imaging and Reporting Data System version 2 in the determination of lesions suspicious for prostate cancer (24). Thus, our results reveal interreader agreement within reported benchmark ranges for diagnostic systems instituted in clinical practice. In addition, adhering to the diagnostic algorithm significantly improved accuracy. Both findings would be desirable in a diagnostic system tailored to support decisions in SRM management.

Nevertheless, opportunities for further refinements in the diagnostic performance of the diagnostic algorithm are obvious. Our reviewer performance was particularly suboptimal for benign renal lesions and chromophobe RCC. Furthermore, a limitation in our study is the small number of chromophobe RCCs and benign tumors. Sensitivities for chromophobe RCC, oncocytoma, and minimal fat AML varied substantially among reviewers, with fair to moderate interreader agreement. In contrast, specificities were consistently high for all three diagnoses (>90% for all reviewers). Schieda et al (25) used a quantitative method and found a sensitivity of 60% and specificity of 97% for the combination of T2W and chemical shift imaging in the diagnosis of minimal fat AML, which is comparable to our results using a qualitative and semiquantitative approach. These authors also reported a sensitivity of 100% and specificity of 89% when combining quantitative T2W and dynamic contrast-enhanced MR imaging (25); this contrasts with our data, which did not show the degree of contrast enhancement or the enhancement characteristics at contrast-enhanced multiphasic MR imaging to be significant predictors of minimal fat AML histology. Since this algorithm had been used for teaching purposes at the authors' institution for some time, it is plausible that some AMLs without visible fat were prospectively recognized and confirmed at percutaneous biopsy. This could result in decreased diagnostic performance of the algorithm in this study cohort, which included only patients who had undergone surgical resection.

Rosenkrantz et al (16) reported that qualitative MR imaging features could not be used to reliably differentiate oncocytomas from chromophobe RCCs. Cornelis et al (13) showed that a quantitative approach could be used to distinguish oncocytomas from chromophobe RCCs with low sensitivity (25%) and high specificity (100%), whereas oncocytomas could be differentiated from clear cell RCC with high sensitivity (100%) and high specificity (94%). While some differences might arise from different patient selection criteria (eg, T1a vs any size tumor), further studies are needed to better understand the discrepancies in the reported diagnostic accuracy for these histopathologic diagnoses.

Some other study limitations must be acknowledged. First, the MR imaging acquisition protocol varied considerably due to the inclusion of examinations performed at different medical centers. The lack of a single optimized protocol might have contributed to the lower diagnostic accuracy. A prereview by a body MR imaging fellowship-trained radiologist was conducted to ensure sufficient image quality and adequacy of image acquisition parameters for all MR studies entered in the final cohort. Second, diffusion-weighted imaging was not included, as it was not consistently performed in all patients. Although there is no strong evidence for the use of diffusion-weighted imaging in the differentiation of histologic subtypes of RCC (26), diffusion-weighted imaging may be used as an ancillary input to differentiate among some renal masses (27). Third, although conclusions on the diagnostic performance could have been influenced by the lack of strict adherence to the algorithm, our results suggest that accuracy was indeed higher when reviewers complied with the algorithm versus when they overruled it; therefore, a trend caused by such factors would likely result in underestimation rather than overestimation of the algorithm performance. Finally, our conclusions may be limited due to the retrospective nature of the study and the prevalence of benign lesions (17%), which is in the lower end of the reported prevalence for T1a disease (16%–30%) (3,28,29). Thus, our cohort may not reflect the actual prevalence of benign disease in patients with SRMs at different centers.

In summary, the proposed standardized diagnostic approach based on multiple MR imaging features helps identify clear cell RCC and papillary RCC with respective accuracies of 81% (88 of 109) and 91% (99 of 109) while achieving moderate to substantial agreement among a large number of radiologists with diverse clinical experience. Further refinements in the algorithm and standardization of the MR imaging protocol may be necessary to improve the diagnostic performance in the characterization of SRMs.

Acknowledgment: The authors thank Pam Curry, MA, for the graphic work in Figure 2. The authors would like to acknowledge UT Southwestern Academic Information Systems for providing data management resources used in this study (Research Electronic Data Capture–REDCap), which were supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under award UL1TR001 105 (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).

Disclosures of Conflicts of Interest: F.U.K. disclosed no relevant relationships. **N.E.C.** disclosed no relevant relationships. **Y.X.** disclosed no relevant relationships. **D.F.P.** disclosed no relevant relationships. **D.N.C.** disclosed no relevant relationships. **A.d.D.L.** disclosed no relevant relationships. **G.K.** disclosed no relevant relationships. **J.R.L.** disclosed no relevant relationships. **T.Y.** disclosed no relevant rela-

tionships. **A.H.L.** disclosed no relevant relationships. **N.K.** disclosed no relevant relationships. **E.K.** disclosed no relevant relationships. **J.A.C.** disclosed no relevant relationships. **I.P.** disclosed no relevant relationships.

References

1. Jayson M, Sanders H. Increased incidence of serendipitously discovered renal cell carcinoma. Urology 1998;51(2):203–205.
2. Hollingsworth JM, Miller DC, Daignault S, Hollenbeck BK. Rising incidence of small renal masses: a need to reassess treatment effect. J Natl Cancer Inst 2006;98(18):1331–1334.
3. Kutikov A, Fossett LK, Ramchandani P, et al. Incidence of benign pathologic findings at partial nephrectomy for solitary renal mass presumed to be renal cell carcinoma on preoperative imaging. Urology 2006;68(4):737–740.
4. Rosales JC, Haramis G, Moreno J, et al. Active surveillance for renal cortical neoplasms. J Urol 2010;183(5):1698–1702.
5. Chawla SN, Crispen PL, Hanlon AL, Greenberg RE, Chen DY, Uzzo RG. The natural history of observed enhancing renal masses: meta-analysis and review of the world literature. J Urol 2006;175(2):425–431.
6. Halverson SJ, Kunju LP, Bhalla R, et al. Accuracy of determining small renal mass management with risk stratified biopsies: confirmation by final pathology. J Urol 2013;189(2):441–446.
7. Wang R, Wolf JS Jr, Wood DP Jr, Higgins EJ, Hafez KS. Accuracy of percutaneous core biopsy in management of small renal masses. Urology 2009;73(3):586–590; discussion 590–591.
8. Vasudevan A, Davies RJ, Shannon BA, Cohen RJ. Incidental renal tumours: the frequency of benign lesions and the role of preoperative core biopsy. BJU Int 2006;97(5):946–949.
9. Tomaszewski JJ, Uzzo RG, Smaldone MC. Heterogeneity and renal mass biopsy: a review of its role and reliability. Cancer Biol Med 2014;11(3):162–172.
10. Oliva MR, Glickman JN, Zou KH, et al. Renal cell carcinoma: t1 and t2 signal intensity characteristics of papillary and clear cell types correlated with pathology. AJR Am J Roentgenol 2009;192(6):1524–1530.
11. Hindman N, Ngo L, Genega EM, et al. Angiomyolipoma with minimal fat: can it be differentiated from clear cell renal cell carcinoma by using standard MR techniques? Radiology 2012;265(2):468–477.
12. Sun MR, Ngo L, Genega EM, et al. Renal cell carcinoma: dynamic contrast-enhanced MR imaging for differentiation of tumor subtypes—correlation with pathologic findings. Radiology 2009;250(3):793–802.
13. Cornelis F, Tricaud E, Lasserre AS, et al. Routinely performed multiparametric magnetic resonance imaging helps to differentiate common subtypes of renal tumours. Eur Radiol 2014;24(5):1068–1080.
14. 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. J Urol 2017;198(4):780–786.
15. Kim JI, Cho JY, Moon KC, Lee HJ, Kim SH. Segmental enhancement inversion at biphasic multidetector CT: characteristic finding of small renal oncocytoma. Radiology 2009;252(2):441–448.
16. Rosenkrantz AB, Hindman N, Fitzgerald EF, Niver BE, Melamed J, Babb JS. MRI features of renal oncocytoma and chromophobe renal cell carcinoma. AJR Am J Roentgenol 2010;195(6):W421–W427.
17. Kay FU, Pedrosa I. Imaging of solid renal masses. Radiol Clin North Am 2017;55(2):243–258.
18. Sasiwimonphan K, Takahashi N, Leibovich BC, Carter RE, Atwell TD, Kawashima A. Small (<4 cm) renal mass: differentiation of angiomyolipoma without visible fat from renal cell carcinoma utilizing MR imaging. Radiology 2012;263(1):160–168.
19. Lopez-Beltran A, Carrasco JC, Cheng L, Scarpelli M, Kirkali Z, Montironi R. 2009 update on the classification of renal epithelial tumors in adults. Int J Urol 2009;16(5):432–443.
20. Lopez-Beltran A, Scarpelli M, Montironi R, Kirkali Z. 2004 WHO classification of the renal tumors of the adults. Eur Urol 2006;49(5):798–805.
21. Pedrosa I, Chou MT, Ngo L, et al. MR classification of renal masses with pathologic correlation. Eur Radiol 2007;18(2):365–375.
22. Roy C, Sauer B, Lindner V, Lang H, Saussine C, Jacqmin D. MR imaging of papillary renal neoplasms: potential application for characterization of small renal masses. Eur Radiol 2007;17(1):193–200.
23. Kerlikowske K, Grady D, Barclay J, et al. Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System. J Natl Cancer Inst 1998;90(23):1801–1809.
24. Rosenkrantz AB, Ginocchio LA, Cornfeld D, et al. Interobserver reproducibility of the PI-RADS version 2 lexicon: a multicenter study of six experienced prostate radiologists. Radiology 2016;280(3):793–804.
25. Schieda N, Dilauro M, Moosavi B, et al. MRI evaluation of small (<4 cm) solid renal masses: multivariate modeling improves diagnostic accuracy for angiomyolipoma without visible fat compared to univariate analysis. Eur Radiol 2016;26(7):2242–2251.
26. 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(2):317–324.
27. Lassel EA, Rao R, Schwenke C, Schoenberg SO, Michaely HJ. Diffusion-weighted imaging of focal renal lesions: a meta-analysis. Eur Radiol 2014;24(1):241–249.
28. Johnson DC, Vukina J, Smith AB, et al. Preoperatively misclassified, surgically removed benign renal masses: a systematic review of surgical series and United States population level burden estimate. J Urol 2015;193(1):30–35.
29. 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(6 Pt 1):2217–2220.