Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

4/20/2012

Caroline Tai Date

Fuhrman Grade is Associated with Radiological Features in Patients with Renal Cell Carcinoma

By

Caroline Tai

Master of Public Health

Epidemiology

Michael Goodman, MD, MPH

Faculty Thesis Advisor

Fuhrman Grade is Associated with Radiological Features in Patients with Renal Cell Carcinoma

By

Caroline Tai

B.S., University of California, San Diego

2008

Faculty Thesis Advisor: Michael Goodman, MD, MPH

An abstract of

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health in Epidemiology

2012

Abstract

Fuhrman Grade is Associated with Radiological Features in Patients with Renal Cell Carcinoma

By Caroline Tai

Background
Tumor grade is an important determinant of RCC prognosis that can influence treatment decisions; it is typically measured using the Fuhrman grading scale, originally introduced in 1982. Identifying factors that could be used to predict Fuhrman Grade (FG) without biopsy prior to surgery would be valuable and would avoid the challenges involved with renal mass biopsies.

Objective
We aim to use these data to evaluate the association between radiological feature scores and FG, while considering other clinical variables.

Methods
The present study is based on 171 renal masses from 171 patients at Emory University between 2006 and 2010. Fuhrman grade was dichotomized into a two-tiered grading system, low (FG I and II) versus high (FG III and IV) since adjacent tumor grades share similar prognosis. Radiological features for each patient were evaluated as score sums calculated by adding up the scores given by four readers. The range across the scores for each kidney mass was also calculated as a measure of agreement or disagreement among the readers.

Results
In unadjusted analyses, Overall Aggressiveness Rating (OAR) ($p<0.0001$), Contour ($p<0.0001$), Venous Invasion ($p=0.0076$), Regional Lymph Node Involvement ($p<0.0001$), Tumor Necrosis ($p=0.0004$), Tumor Consistency ($p=0.0031$), Hilar Status ($p=0.0022$), and Collateral Vascularity ($p<0.0001$) were all significantly associated with high tumor grade. OAR score sum was significantly associated with high FG ($p=0.0029$) while controlling for age, sex, tumor size, and disagreement (OAR score range).

Conclusions
This study confirmed the association between FG and radiological features such as OAR, indicating that there may be predictive value in radiological features.

Fuhrman Grade is Associated with Radiological Features in Patients with Renal Cell Carcinoma

By

Caroline Tai

B.S., University of California, San Diego

2008

Faculty Thesis Advisor: Michael Goodman, MD, MPH

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology

2012

Acknowledgements

I would like to thank my thesis advisor, Dr. Michael Goodman for his guidance and support that made this thesis possible. I am indebted to Dr. Viraj Master and Dr. Emmanuel Reyes for sharing their subject matter expertise and knowledge of this dataset. And finally, I am grateful for the tireless efforts of Ruth Westby who provided additional data for this project.

Table of Contents

Background…… Page 1

Methods…….. Page 4

Results……..…… Page 10

Discussion……...…. Page 14

References. ……..…….…. Page 17

Tables …….……..…….. Page 21

Figures ……..……...…… Page 30

Background

Incidence of kidney and renal pelvis cancers has been increasing each year for the past few decades. In 2011, 60,920 cases of this malignancy were estimated to occur in the United States, with an estimated 13,120 deaths (23). Kidney cancers are a heterogeneous group that includes renal cell carcinoma (RCC), renal pelvis carcinoma, and Wilms tumor. The most common of those is RCC, which arises in the renal parenchyma and accounts for 92% of all incident kidney cancer cases (23). Additionally, within RCC there are several histologic subtypes; the five most common include clear cell, papillary, chromophobe, hereditary, and multilocular cystic (21). Most kidney cancers including RCC are treated with surgical resection that involves removal of a part or the entire kidney (partial or radical nephrectomy). The economic burden of this malignancy has been estimated to be $4.4 billion based on SEER-Medicare data (16). Currently, screening is not conducted for kidney cancer in the general population and it is typically found incidentally on imaging by computerized tomography (CT) or ultrasound conducted for other indications such as abdominal pain, gynecologic issues, or gastrointestinal problems (18). After incidental discovery, additional CT scans are taken to characterize renal tumors in more detail.

Tumor grade is an important determinant of RCC prognosis that can influence treatment decisions, it is usually measured using the Fuhrman grading

scale, originally introduced in 1982 (6, 9). Fuhrman grade (FG) has been shown in numerous studies to be one of the most significant predictors of survival and is currently the most widely accepted grading system for RCC in clinical practice (25). FG is organized into four grades (I, II, III, and IV) and is based on nuclear characteristics including nuclear size, nuclear pleomporphism, and nucleolar prominence (20). Predicted 5-year survival for RCC ranges from 91% to 32%, for FG I and FG IV respectively (3). After imaging by CT or ultrasound, many patients undergo renal mass biopsy (RMB) to obtain tissue samples for histologic processing which is then used to determine FG before considering treatment options and prognosis. However due to concerns about biopsy failure rates, reported to be as high as 37% for fine needle biopsy or because of the likelihood of indeterminate pathology, shown to be as high as 36%, only 30-67% of patients undergo RMB (13,17,24). Alternatively FG may be determined post-operatively using tissue samples obtained from nephrectomy; in these cases, FG is not considered before surgery is performed. FG and Classification of Malignant Tumors (TNM) staging are the strongest independent prognostic factors for localized RCC, thus FG is important to consider prior to surgery.

Identifying factors that could be used to predict FG without biopsy prior to surgery would be valuable and would avoid the challenges involved with RMB. Jeldres et al. 2009 developed a prediction model for FG that incorporated age, gender, tumor size and symptom classification (asymptomatic, local, and

systemic) that yielded an ROC curve with 58.3% area under the curve. Other areas of oncology use imaging studies characteristics to predict histologic features of the tumors, including grade. For example, Khalid et al. 2012 proposed incorporating three imaging characteristics, contrast enhancement, calcification, and apparent diffusion coefficient, into a model for tumor grade prediction in oligodendrogliomas of the brain. Only one study has attempted to develop a FG prediction model using anatomic features from imaging; Kutikov et al 2011 developed a nomogram which incorporated radiographic features using the RENAL Nephrometry Score (**R**adius, **E**xophytic properties, **N**earness of tumor to collecting system or sinus, **A**nterior/posterior, **L**ocation relative to the polar lines) in the prediction of high FG (12). This nomogram was developed in a patient population from the Fox Chase Cancer Center in Philadelphia, PA, and was later validated in a Chinese patient cohort from Fudan University Shanghai Cancer Center by Wang et al 2012.

The present study is based on our experience with 171 patients evaluated for renal masses at Emory University between 2006 and 2010. We aim to use these data to evaluate the association between radiological feature scores and FG, while considering other clinical variables.

Methods

Patient Cohort

Data collection and analysis protocols were approved by the Emory Institutional Review Board for human subjects' research. Pathology and laboratory information on patients who have undergone radical or partial nephrectomy was obtained from the patient records maintained by the Department of Urology at Emory Healthcare in Atlanta, GA. Each renal mass was viewed as a separate observation and was included in the current analysis based on the availability of a pre-operative contrast-enhanced CT imaging study, renal masses with only non-contrast CTs were excluded. Each renal mass was considered eligible for inclusion only if the radiographic assessment was accompanied with a subsequent post-operative pathology report with a documented FG. This yielded 196 patients with 205 renal masses. A total of 25 renal masses had not been assigned a grade and were excluded. For instances in which multiple renal masses were identified for a single patient, only the first incident renal mass was included. The final analytic dataset included 171 renal masses from 171 patients.

Clinical Variables

Clinically relevant variables abstracted from medical records included age at surgery, race, gender, smoking status (yes/no), statin use (yes/no), body mass index (BMI), blood pressure, number of medications used prior to surgery,

tumor histology subtype, and mass size and laterality. Laboratory measures included serum creatinine, blood urea nitrogen (BUN), concentrations of hemoglobin, albumin and C-reactive protein (CRP), white blood cell and platelet counts, and erythrocyte sedimentation rate (ESR). When multiple laboratory measures were available, the most recent measurement taken prior to the surgery date was recorded. Tumor grade was assigned by the attending physician using criteria described in Fuhrman et al 1982.

Radiological Features

Following a literature review and discussion, a research committee of three urologists (readers B, C, D) and a radiologist (reader A) experienced in urogenital imaging selected eleven radiological features that could be considered predictive of biological kidney tumor behavior (Table 1). The same clinicians also served as the readers, each assigning a score for every radiological feature after viewing the images on the same size screen. CT imaging was evaluated independently for each renal mass and all readers were blinded to each patient's clinical history and pathological findings.

Overall Aggressiveness Rating (OAR) was the readers' initial perception of aggressiveness. The scores for OAR were assigned as follows: 0 indicated a non-aggressive mass that is likely to be benign, 1 indicated a minimally aggressive mass believed to be malignant, and 2 indicated the mass was aggressive and prognosis for survival would be poor. Contrast enhancement was

assigned 0 if the renal mass was less dense than the parenchyma, 1 if the renal mass had the same density as the parenchyma, and 2 if the renal mass was denser than the parenchyma. Contour described the borders of the mass where 0 indicated a well-circumscribed or round shape, 1 indicated that readers had some difficulty defining the border or the mass appeared lobulated, and 2 indicated that readers were unable to define borders or the mass was ill-marginated.

Venous invasion was assigned 0 if there was no venous involvement, 1 for segmental or renal vein involvement, and 2 for inferior vena cava involvement. Regional Lymph node involvement evaluated from both the axial and coronal views was assigned 0 if there was no lymph node involvement, 1 if the involved lymph nodes were likely to be benign because regional nodes were less than 1 cm or there was mild enhancement of involved nodes or if relatively few lymph nodes were involved, and 2 if involved nodes were likely malignant because regional nodes were greater than 1 cm or there was significant enhancement of involved nodes or there was involvement of a relatively significant number of lymph nodes. Perilesional fat stranding which compared the affected kidney to the healthy kidney was assigned 0 if there was no stranding or that the stranding was similar in both kidneys, 1 if there was moderate disproportionate stranding, and 2 if there was severe disproportionate stranding.

Tumor necrosis was assigned 0 if there was no necrosis, 1 for mild necrosis defined as less than 25% of the tumor volume , and 2 for severe necrosis defined as greater than 25% of the tumor volume. Tumor consistency was the only binary radiological feature where a mass was identified as cystic or multicystic (0) or solid (1). The location of the tumor mass was assigned 0 if exophytic defined as <25% of the mass located within the kidney, 1 if mesophytic defined as between 25 -75% of the mass located within the kidney, and 2 if endophytic defined as >75% of the mass located within the kidney. Hilar status was assigned 0 if non-hilar, 1 if near the hilum defined as < 25% of the tumor mass located within the hilar box, and 2 if the mass was hilar defined as > 25% of the tumor located within the hilar box. The hilar box is shown in Figure 1. Collateral vascularity was described as severe (2), mild (1), or no (0) collateral vasculature development. Tumor size was determined by the largest tumor diameter (cm) on CT imaging.

Statistical Methods

Fuhrman grade was dichotomized as low (FG I and II) versus high (FG III and IV) since adjacent tumor grades share similar prognosis and the two-tiered grading system has been used in previous studies (1, 16, 2, 6). Age was divided into 20-year intervals: 30-49, 50-69, and 80-89. Race/ethnicity was categorized as non-Hispanic White (NHW) versus other. Unadjusted analyses compared low-grade to high-grade renal masses with respect to patient's age at surgery, race,

gender, smoking status, statin use, BMI, number of medications used, blood pressure, and tumor laterality, histology, and size. These unadjusted comparisons were accompanied by chi-square tests for categorical variables and Student's t-tests for continuous variables. Additionally Student's t-tests were used to compare the distributions of laboratory measures between the dichotomized grade categories.

Inter-rater agreement for each radiological parameter under study was evaluated using a mean weighted kappa (κ) statistic (4, 8). A κ statistic was calculated for each possible pair of raters and the mean was taken for all pairs; this method is described in Conger 1980. Interpretation of the κ statistic was based on cutoffs reported in previous studies where κ values of 0.00–0.20, 0.21–0.45, 0.46–0.75, 0.76–0.99 and 1.0 were considered as showing fair, moderate, substantial, almost perfect and perfect agreement, respectively (16, 19, 1). To investigate whether a single reader affected the overall score, each reader's assigned points were removed, and weighted kappa statistics were recalculated with the remaining three readers; this was repeated for every radiological feature. A Spearman rank-correlation matrix was constructed to describe the relation between paired radiological features.

To assess the relation between radiological parameters and tumor grade, the sum of four scores for each radiological feature was calculated to obtain a single value. Values for these score sums ranged from 0 (when all four readers

assigned 0 points) to 8 (when all readers assigned 2 points). This 0-8 range applied to all radiographic parameters with the exception of the binary (0 vs. 1) measure, Tumor Consistency, which ranged from 0 to 4. It is important to note that a score sum in the middle of the possible range could represent different combinations of individual scores; for example, a score of 4 could be the results of each reader assigning 1 point or 1, 1, 1, 1 versus 1,1,0,2 versus 0, 0, 2, 2. For this reason the range across the four scores for each kidney mass was also calculated for each parameter as a measure of agreement among the readers.

Once each kidney mass underwent evaluation and scoring by all four readers the distributions of summary scores were compared in high and low grade tumors using chi-square, Fisher's exact or Fisher Freeman Halton tests, as appropriate. Multivariate logistic regression analysis included age, gender, tumor size, and OAR (score sum and range). The variables age, gender and tumor size have been consistently reported in previous studies to be associated with high FG (15, 10, 12, 26). The addition of OAR score range is to control for disagreement among readers. Lane et al. found smoking status to be associated with FG, therefore this variable was also added to the model. The likelihood ratio χ^2 statistic was calculated for all models. All statistical analyses were performed with SAS statistical software, version 9.3 (SAS institute Inc, Cary, NC). All tests were defined as significant based on the two-sided α-error cutoff of 0.05.

Results

Descriptive statistics and unadjusted analyses comparing high- and low-grade tumors are shown in Table 1. About half of the 171 tumors (n=87) were considered low-grade; of those, 7 tumors were assigned FG I and 80 tumors were assigned FG II. Among the 89 high-grade tumors, FG III and FG IV were assigned to 65 and 19 tumors, respectively.

Overall, more patients fell in the age group 50-69 yrs (55.6%) than in 30-49 yrs (19.3%) or 70-89 yrs (25.1%) and this distribution was similar between low and high grade. Out of 157 patients, there were 60 patients of white race with high grade and 56 patients with low grade but this difference did not reach statistical significance. The majority of patients classified as other race were of black race (n=18 with low grade, 16 with high grade) the other races included Asian (n=2 with low grade, 1 with high grade), Hispanic (n=1 with low grade, 0 with high grade) and other (n=2 in low grade, 1 in high grade). In this cohort, 111 patients were male (64.9%), 80 were smokers (46.8%) and the distribution of clear cell (74.9%), papillary (17.0%) and other (8.2%) subtypes was similar between low and high grade (p=0.7158). Patients classified as other subtype included 5 with chromophobe subtype, 1 of multilocular cystic subtype, 1 of sarcomatoid, 3 of clear-cell papillary subtype, and 4 renal masses had subtypes not otherwise specified. Mean BMI was similar in low grade patients (30.0 kg/m^2) compared to

high grade patients (28.6 kg/m^2). Mean systolic blood pressure was similar in both groups (133 mmHg in low grade, 131 mmHg in high grade) while mean diastolic blood pressure was slightly elevated in low grade (76 mmHg) compared to high grade patients (73 mmHg) but this was not a statistically significant difference. For the univariate analyses in Table 1, only mean tumor size was significantly associated with high grade ($p<0.001$). Laboratory measures are summarized in Table 2. Mean hemoglobin ($p=0.0003$), albumin ($p=0.0243$), CRP ($p=0.0020$), and ESR ($p=0.0031$) were all associated with high tumor grade in univariate analyses. There were notable numbers of patients with missing values for CRP (n=49) and ESR (n=50) since these are not part of the standard laboratory panel.

Table 3 summarizes the score sums for each radiological feature. In unadjusted analyses, OAR ($p<0.0001$), Contour ($p<0.0001$), Venous Invasion ($p=0.0076$), Regional Lymph Node Involvement ($p<0.0001$), Tumor Necrosis ($p=0.0004$), Tumor Consistency ($p=0.0031$), Hilar Status ($p=0.0022$), and Collateral Vascularity ($p<0.0001$) were all significantly associated with high tumor grade. The ranges of the radiological scores are shown in Table 4 where perfect agreement is quantified by a score range of 0. For OAR, readers exhibited greater agreement in high FG kidney masses (35.7% with score range 0-3) compared to low FG (20.7% with score range 0-3), $p=0.0499$. This was also the case for Tumor Consistency where agreement was higher in high FG than low

FG, p=0.0073. Conversely, Venous Invasion (p=0.0061), Regional Lymph Node Involvement (p=0.0002), Perilesional Fat Stranding (p=0.0043) and Collateral Vascularity (p=0.0424) showed higher agreement in low FG than high FG.

The mean weighted kappa statistics for the radiological features in Table 5 indicate moderate to substantial inter-rater agreement. For OAR agreement was moderate when all four readers were compared (mean κ=0.43), this was largely unaffected by the removal of any reader, mean κ ranged from 0.38 to 0.44 for each subset of readers. The highest overall agreement was achieved for Venous Invasion (mean κ=0.63) followed by Tumor Necrosis (mean κ=0.61), Collateral Vascularity (mean κ=0.59) and Hilar status (mean κ=0.58). The subgroup of readers B, C, D consists of only urologists and among these readers, Tumor Necrosis showed the highest agreement (mean κ=66). Generally, the inclusion of a radiologist in a group of urologists did not greatly change the κ statistic. For Perilesional Fat Stranding weighted kappa statistics could not be calculated for pairs that included reader C because this reader did not assign a score of 2 to any patient. The spearman rank correlation matrix in Table 6 indicated that OAR was significantly correlated ($p<0.0001$) with all other radiological features, spearman correlation values ranged from 0.3216 to 0.7717.

Table 7 describes the multivariate logistic regression model in which model 1 included OAR score sum and range, age, sex, and tumor size. OAR score sum was significantly associated with high FG (p=0.0029) while controlling for

age, sex, tumor size, and disagreement (OAR score range). Adding smoking status (model 2) to the first model did not greatly change the likelihood ratio χ^2 statistic nor the association between OAR and FG. Overall model 2 was similar to model 1, thus it may not be necessary to control for smoking status.

Discussion

The heterogeneity of kidney cancers makes treatment decisions and prognostic determination difficult. The two main predictors of kidney cancer clinical behavior include TNM staging, and FG; however, the use of FG has been subject to criticisms due to variable inter-rater agreement and variable performance across histologic subtypes other than clear cell (25). FG assigned from RMB is vulnerable to false-negative and false-positive results; the former being most disconcerting since it may leave a renal mass untreated with a potential for metastasis. Reported accuracy estimates for predicting high and low grade disease using fine-needle and core biopsy have been as low as 28% and 76%, respectively, when compared to nephrectomy-derived FG (22). The uncertainty of biopsy derived grade has complicated the use of FG as a prognostic predictor in clinical practice. Furthermore, the value of RMB for determining FG has been limited by insufficient tissue sampling resulting in indiscriminate histology, which can occur in 21% of cases (17). This has spurred the development of FG prediction models using other clinical variables that may be readily available from patient records. This study confirmed the association between FG and radiological features such as OAR, indicating that there may be predictive value in radiological features.

A major strength of this study is the relatively equal proportion of high and low FG patients. This allows for greater discernment between high and low grade disease. Previous studies have had a lower proportion, 38.3% to 45%, of their patient cohort with high FG (12, 26). A model developed in a population of mostly low FG may need to be further validated for patient populations with greater proportions of high FG.

Following the nomogram developed by Kutikov et al. and validated by Wang et al., this study has also identified other radiological features that may be considered to predict FG and prognosis. These potential predictors were represented in multivariate models by the addition of OAR which showed at least moderate inter-rater agreement and significantly correlated with all radiological features presented in this study. Furthermore the final model was adjusted for disagreement in OAR scores for each kidney mass among readers, a component that has not been incorporated into previous models.

The limitations of this study include relatively modest inter-rater agreement for OAR. The reliability of score assignment may need to be further tested before implementation in clinical practice. Since the research committee also served as the raters, it is unclear if other clinicians will interpret the definitions for each radiological feature as intended.

Delahunt et al. indicated that FG prediction models may perform variably in different histologic subtype. All kidney masses in these analyses were treated

without regard to differential subtype. However most masses were of clear cell (74.9%) and papillary (17.0%) subtype and only 8.2% belonged to other subtypes. There were too few observations to evaluate the data by subtype but validation of this model in a larger population would allow for these analyses. Validation of this model is also needed since it was developed using data from a single center and generalizability of these findings to other populations may be limited.

Future efforts should be directed at analyzing the other radiological features for predictive and prognostic value. It may also be useful to consider the addition of laboratory measures such as CRP and albumin to a prognostic model as proposed by Lamb et al 2006. With the large number of variables available in this dataset, the models presented by Kutikov et al and Lane et al can be tested in this patient cohort and compared to a model incorporating the radiological features presented in this study. This study is the first to identify 11 radiological features that could later be used in prognostic and FG prediction models. Further investigation aimed at building such predictive models will enhance the utility of these radiographic features.

References

1. Al-Aynati M, Chen V, Salama S, Shuhaibar H, Treleaven D, Vincic L. Interobserver and intraobserver variability using the Fuhrman grading system for renal cell carcinoma. *Arch Pathol Lab Med*. 2003 May;127(5):593-6.
2. Bektas S, Bahadir B, Kandemir NO, Barut F, Gul AE, Ozdamar SO. Intraobserver and interobserver variability of Fuhrman and modified Fuhrman grading systems for conventional renal cell carcinoma. *Kaohsiung J Med Sci*. 2009 Nov;25(11):596-600.
3. Chin AI, Lam JS, Figlin RA, Belldegrun AS. Surveillance strategies for renal cell carcinoma patients following nephrectomy. *Rev Urol.* 2006 Winter;8(1):1-7.
4. Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin.* 1968;70:213-220
5. Conger AJ. Integration and Generalization of Kappas for Multiple Raters. *Psychological Bulletin*. 1980. Vol. 88, No. 2, 322-328.
6. Delahunt, Brett. Advances and controversies in grading and staging of renal cell carcinoma. *Modern Pathology.* 2009;22:S24-S36.
7. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988; 44:837–845.

8. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychological Bulletin*. 1971;76:378-382.
9. Fuhrman SA, Lasky LC and Limas C: Prognostic significance of morphologic parameters in renal cell carcinoma. *Am J Surg Pathol*. 1982; 6: 655.
10. Jeldres C, Sun M, Liberman D, Lughezzani G, de la Taille A, Tostain J, Valeri A, Cindolo L, Ficarra V, Artibani W, Zigeuner R, Mejean A, Descotes JL, Lechevallier E, Mulders PF, Perrotte P, Patard JJ, Karakiewicz PI.Can renal mass biopsy assessment of tumor grade be safely substituted for by a predictive model? *J Urol*. 2009 Dec;182(6):2585-9.
11. Khalid L, Carone M, Dumrongpisutikul N, Intrapiromkul J, Bonekamp D, Barker PB, Yousem DM. Imaging Characteristics of Oligodendrogliomas That Predict Grade. *AJNR Am J Neuroradiol*. 2012 Jan 19.
12. Kutikov A, Uzzo RG.The R.E.N.A.L. nephrometry score: a comprehensive standardized system for quantitating renal tumor size, location and depth. *J Urol*. 2009 Sep;182(3):844-53. Epub 2009 Jul 17.
13. Laguna MP, Kümmerlin I, Rioja J, de la Rosette JJ. Biopsy of a renal mass: where are we now? *Curr Opin Urol*. 2009 Sep;19(5):447-53.
14. Lamb GW, McMillan DC, Ramsey S, Aitchison M. The relationship between the preoperative systemic inflammatory response and cancer-specific survival in patients undergoing potentially curative resection for renal clear cell cancer.*Br J Cancer*. 2006 Mar 27;94(6):781-4.

15. Lane BR, Babineau D, Kattan MW, Novick AC, Gill IS, Zhou M, Weight CJ, Campbell SC.A preoperative prognostic nomogram for solid enhancing renal tumors 7 cm or less amenable to partial nephrectomy. *J Urol*. 2007 Aug;178(2):429-34. Epub 2007 Jun 11.
16. Lang H, Lindner V, de Fromont M, Molinié V, Letourneux H, Meyer N, Martin M, Jacqmin D. Multicenter determination of optimal interobserver agreement using the Fuhrman grading system for renal cell carcinoma: Assessment of 241 patients with > 15-year follow-up. *Cancer*. 2005 Feb 1;103(3):625-9
17. Lechevallier E, André M, Barriol D, Daniel L, Eghazarian C, De Fromont M, Rossi D, Coulange C. Fine-needle percutaneous biopsy of renal masses with helical CT guidance. *Radiology*. 2000 Aug;216(2):506-10.
18. McCauley LR, Beckham CJ, Hunter TB, Nguyen MM. Gender and renal cancer: do variations in clinical presentation and imaging patterns explain observed differences between males and females? *Urology*. 2010 Sep;76(3):536-40. Epub 2010 Feb 6.
19. Munoz S, Bangdiwala S. Interpretation of kappa and B statistics measures of agreement. *J Appl Stat*. 1997;24:105–111.
20. Novara G, Martignoni G, Artibani W, Ficarra V. Grading Systems in renal cell carcinoma. (2007) *J Urol*. Feb;177(2):430-6.
21. Prasad SR, Humphrey PA, Catena JR, Narra VR, Srigley JR, Cortez AD, Dalrymple NC, Chintapalli KN. Common and uncommon histologic

subtypes of renal cell carcinoma: imaging spectrum with pathologic correlation. *Radiographics*. 2006 Nov-Dec;26(6):1795-806; discussion 1806-10.

22. Schmidbauer J, Remzi M, Memarsadeghi M, Haitel A, Klingler HC, Katzenbeisser D, Wiener H, Marberger M.Diagnostic accuracy of computed tomography-guided percutaneous biopsy of renal masses. *Eur Urol.* 2008 May;53(5):1003-11. Epub 2007 Nov 26.
23. Siegel, R., Ward, E., Brawley, O. and Jemal, A. (2011), Cancer statistics, 2011. *CA: A Cancer Journal for Clinicians*, 61: 212–236. doi: 10.3322/caac.20121
24. Tan HJ, Jacobs BL, Hafez KS, Montgomery JS, Weizer AZ, Wood DP Jr, Miller DC, Wolf JS Jr. Understanding the role of percutaneous biopsy in the management of patients with a small renal mass. *Urology*. 2012 Feb;79(2):372-7.
25. Volpe A, Patard JJ. Prognostic factors in renal cell carcinoma. *World J Urol*. 2010 Jun;28(3):319-27. Epub 2010 Apr 3. Review.
26. Wang HK, Zhu Y, Yao XD, Zhang SL, Dai B, Zhang HL, Shen YJ, Wang CF, Ye DW. External Validation of a Nomogram Using RENAL Nephrometry Score to Predict High Grade Renal Cell Carcinoma. *J Urol*. 2012 Mar 14.

Tables

Table 1. Patient and renal mass characteristics.

Patient Characteristics	**Overall (N = 171)**	**FG I/II (N = 87)**	**FG III/IV (N = 84)**	***p-value*†**
Age, n (%)				
30-49 yrs	33 (19.3%)	21 (24.1%)	12 (14.3%)	
50-69 yrs	95 (55.6%)	48 (55.2%)	47 (56.0%)	
70-89 yrs	43 (25.1%)	18 (20.7%)	25 (29.8%)	0.1692
Race, n (%)				
White	116 (67.8%)	56 (64.4%)	60 (71.4%)	
Other	41 (24.0%)	23 (26.4%)	18 (21.4%)	0.3892
Missing, n	*14*	*8*	*6*	
Sex (male), n (%)	111 (64.9%)	53 (60.9%)	58 (69.0%)	0.2655
Smoker, n (%)	80 (46.8%)	41 (47.1%)	39 (46.4%)	0.9305
Missing, n	*4*	*2*	*2*	
Statin use, n (%)	56 (32.7%)	29 (33.3%)	27 (32.1%)	0.9248
Missing, n	*18*	*7*	*11*	
BMI (kg/m^2), mean ± SD	29.3± 6.9	30.0 ± 7.0	28.6 ± 6.8	0.1849
Missing, n	*3*	*2*	*1*	
Medications (n), mean ± SD	5.7 ± 4.0	5.7 ± 4.4	5.7 ± 3.5	0.9935
Missing, n	*1*	*1*	*0*	
Systolic blood pressure (mmHg), mean ± SD	132± 19	133 ± 18	131 ± 21	0.5484
Missing, n	*6*	*3*	*3*	
Diastolic blood pressure (mmHg), mean ± SD	74± 11	76 ± 11	73 ± 11	0.0760
Missing, n	*6*	*3*	*3*	
Renal Mass Characteristics	**Overall (N = 171)**	**FG I/II (N = 87)**	**FG III/IV (N = 84)**	***p-value*†**
Laterality (left), n (%)	82 (48.0%)	41 (47.1%)	41 (48.8%)	0.8257
Tumor Histology, n (%)				
Clear Cell	128 (74.9%)	66 (75.9%)	62 (73.8%)	
Papillary	29 (17.0%)	13 (14.9%)	16 (19.0%)	
Other	14 (8.2%)	8 (9.2%)	6 (7.1%)	0.7158
Tumor Size (cm), mean ± SD	4.89± 2.92	3.88± 2.38	5.92± 3.06	**<0.0001**†
Missing, n	*1*	*1*	*0*	

Values are reported as mean ± standard deviation for continuous variables and as frequency (n) and percentages (%) for categorical variables

† Chi-square tests were used for categorical variables and student's t-test were used for continuous variables (α=0.05)

FG = Fuhrman Grade

Table 2. Laboratory Measures

Blood test results	Overall N = 171	FG I/II N = 87	FG III/IV N = 84	*p-value*†
Creatinine (mg/dL)	1.73 ± 2.37	1.85 ± 2.62	1.60 ± 2.10	0.5020
Blood urea nitrogen (mg/dL)	16.29 ± 8.61	16.01 ± 9.47	16.58 ± 7.67	0.6673
Hemoglobin(g/dL)	13.20 ± 1.82	13.69 ± 1.62	12.69 ± 1.88	**0.0003**†
Missing	*1*	*1*	*0*	
Albumin(g/dL)	16.46 ± 40.12	3.67 ± 0.40	3.48 ± 0.61	**0.0243**†
Missing	*2*	*0*	*2*	
Whole blood cell count (x10^3/µL)	7.73 ± 6.92	7.90 ± 9.26	7.56 ± 3.08	0.7471
Missing	*1*	*1*	*0*	
Platelet count (x 10^3/µL)	257 ± 87	256 ± 66	259 ± 105	0.8168
Missing	*1*	*1*	*0*	
C-reactive protein (mg/dL)	16.46 ± 40.12	5.13 ± 8.78	26.41 ± 52.54	**0.0020**†
Missing	*49*	*30*	*19*	
Erythrocyte sedimentation rate (mm/h)	30.96 ± 29.22	22.91 ± 21.35	38.13 ± 33.31	**0.0031**†
Missing	*50*	*30*	*20*	

Values are reported as mean ± standard deviation
† Student's t-test were used to test for difference in means (α=0.05)
FG = Fuhrman Grade

Table 3. Score Sums for Radiological Features

Radiologic features	Overall N = 171	FG I/II N = 87	FG III/IV N = 84	*p-value*†
Overall Aggressiveness Rating				
Score sum 0-3	64 (35.6%)	43 (49.4%)	21 (25.0%)	
Score sum 4-6	64 (37.4%)	34 (39.1%)	30 (35.7%)	
Score sum 7-8	42 (24.6%)	9 (10.3%)	33 (39.3%)	**<0.0001**†
Missing	*1*	*1*	*0*	
Degree Enhancement				
Score sum 0-3	86 (50.3%)	47 (54.0%)	39 (46.4%)	
Score sum 4-6	69 (40.4%)	32 (36.8%)	37 (44.0%)	
Score sum 7-8	15 (8.8%)	7 (8.0%)	8 (9.5%)	0.5628
Missing	*1*	*1*	*0*	
Contour				
Score sum 0-3	114 (66.7%)	68 (78.2%)	46 (54.8%)	
Score sum 4-6	30 (17.5%)	14 (16.1%)	16 (19.0%)	
Score sum 7-8	26 (15.2%)	4 (4.6%)	22 (26.2%)	**0.0001***
Missing	*1*	*1*	*0*	
Venous Invasion				
Score sum 0-3	154 (90.1%)	83 (95.4%)	71 (84.5%)	
Score sum 4-6	9 (5.3%)	3 (3.4%)	6 (7.1%)	
Score sum 7-8	7 (4.1%)	0	7 (8.3%)	**0.0076***
Missing	1	1	0	
Regional Lymph Node Involvement				
Score sum 0-3	156 (91.2%)	86 (98.9%)	70 (83.3%)	
Score sum 4-6	10 (5.8%)	0	10 (11.9%)	
Score sum 7-8	4 (2.3%)	0	4 (4.8%)	**<0.0001***
Missing	*1*	*1*	*0*	
Perilesional Fat Stranding				
Score sum 0-3	152 (88.9%)	81 (93.1%)	71 (84.5%)	
Score sum 4-6	16 (9.4%)	5 (5.7%)	11 (13.1%)	0.0934
Missing	*3*	*1*	*2*	
Tumor Necrosis				
Score sum 0-3	93 (54.4%)	59 (67.8%)	34 (40.5%)	
Score sum 4-6	45 (26.3%)	19 (21.8%)	26 (31.0%)	
Score sum 7-8	32 (18.7%)	8 (9.2%)	24 (28.6%)	**0.0004**†
Missing	*1*	*1*	*0*	
Tumor Consistency				
Score sum 0-2	9 (5.3%)	9 (10.3%)	0	
Score sum 3-4	161 (94.2%)	77 (88.5%)	84 (100.0%)	**0.0031***
Missing	*1*	*1*	*0*	
Location				
Score sum 0-3	42 (24.6%)	23 (26.4%)	19 (22.6%)	
Score sum 4-6	81 (47.4%)	43 (49.4%)	38 (45.2%)	
Score sum 7-8	47 (27.5%)	20 (23.0%)	27 (32.1%)	0.4255
Missing	*1*	*1*	*0*	

Table 3. Score Sums for Radiological Features (continued)

Radiologic features	Overall N = 171	FG I/II N = 87	FG III/IV N = 84	*p-value*†
Hilar status				
Score sum 0-3	103 (60.2%)	60 (69.0%)	43 (51.2%)	
Score sum 4-6	36 (21.1%)	19 (21.8%)	17 (20.2%)	
Score sum 7-8	31 (18.1%)	7 (8.0%)	24 (28.6%)	**0.0022**†
Missing	*1*	*1*	*0*	
Collateral Vascularity				
Score sum 0-3	130 (76.0%)	77 (88.5%)	53 (63.1%)	
Score sum 4-6	28 (16.4%)	8 (9.2%)	20 (23.8%)	
Score sum 7-8	12 (7.0%)	1 (1.1%)	11 (13.1%)	**0.0001***
Missing	*1*	*1*	*0*	

Values are reported as frequency (n) and percentages (%)
*Fisher Freeman Halton test or Fisher's exact test was used where cells counts < 5 (α=0.05)
† Chi-square tests were used (α=0.05)

Table 4. Score Ranges for Radiological Features

Radiologic features	Overall N = 171	FG I/II N = 87	FG III/IV N = 84	*p-value*†
Overall Aggressiveness Rating				
Score Range=0	48 (28.1%)	18 (20.7%)	30 (35.7%)	
Score Range=1	114 (66.7%)	62 (71.3%)	52 (61.9%)	
Score Range=2	8 (4.7%)	6 (6.9%)	2 (2.4%)	**0.0499***
Missing	*1*	*1*	*0*	
Degree Enhancement				
Score Range=0	65 (38.0%)	37 (42.5%)	28 (33.3%)	
Score Range=1	88 (51.5%)	43 (49.4%)	45 (53.6%)	
Score Range=2	17 (9.9%)	6 (6.9%)	11 (13.1%)	0.2542
Missing	*1*	*1*	*0*	
Contour				
Score Range=0	64 (37.4%)	30 (34.5%)	34 (40.5%)	
Score Range=1	83 (48.5%)	40 (46.0%)	43 (51.2%)	
Score Range=2	23 (13.5%)	16 (18.4%)	7 (8.3%)	0.1454
Missing	*1*	*1*	*0*	
Venous Invasion				
Score Range=0	137 (80.1%)	77 (88.5%)	60 (71.4%)	
Score Range=1	25 (14.6%)	8 (9.2%)	17 (20.2%)	
Score Range=2	8 (4.7%)	1 (1.1%)	7 (8.3%)	**0.0061***
Missing	1	1	0	
Regional Lymph Node Involvement				
Score Range=0	113 (66.1%)	68 (78.2%)	45 (53.6%)	
Score Range=1	45 (26.3%)	17 (19.5%)	28 (33.3%)	
Score Range=2	12 (7.0%)	1 (1.1%)	11 (13.1%)	**0.0002***
Missing	*1*	*1*	*0*	
Perilesional Fat Stranding				
Score Range=0	92 (53.8%)	57 (65.5%)	35 (41.7%)	
Score Range=1	66 (38.6%)	27 (31.0%)	39 (46.4%)	
Score Range=2	10 (5.8%)	2 (2.3%)	8 (9.5%)	**0.0043***
Missing	*3*	*1*	*2*	
Tumor Necrosis				
Score Range=0	85 (49.7%)	43 (49.4%)	42 (50.0%)	
Score Range=1	60 (35.1%)	33 (37.9%)	27 (32.1%)	
Score Range=2	25 (14.6%)	10 (11.5%)	15 (17.9%)	0.4519
Missing	*1*	*1*	*0*	
Tumor Consistency				
Score Range=0	148 (86.5%)	69 (79.3%)	79 (94.0%)	
Score Range=1	22 (12.9%)	17 (19.5%)	5 (6.0%)	**0.0073**†
Missing	*1*	*1*	*0*	
Location				
Score Range=0	75 (43.9%)	38 (43.7%)	37 (44.0%)	
Score Range=1	88 (51.5%)	45 (51.7%)	43 (51.2%)	
Score Range=2	7 (4.1%)	3 (3.4%)	4 (4.8%)	0.9650
Missing	*1*	*1*	*0*	

Table 4. Score Ranges for Radiological Features (continued)

Radiologic features	Overall N = 171		FG I/II N = 87		FG III/IV N = 84		*p-value*†
Hilar status							
Score Range=0	81	(47.4%)	46	(52.9%)	35	(41.7%)	
Score Range=1	64	(37.4%)	28	(32.2%)	36	(42.9%)	
Score Range=2	25	(14.6%)	12	(13.8%)	13	(15.5%)	0.2850
Missing	*1*		*1*		*0*		
Collateral Vascularity							
Score Range=0	94	(55.0%)	55	(63.2%)	39	(46.4%)	
Score Range=1	68	(39.8%)	29	(33.3%)	39	(46.4%)	
Score Range=2	8	(4.7%)	2	(2.3%)	6	(7.1%)	**0.0424***
Missing	*1*		*1*		*0*		

Values are reported as frequency (n) and percentages (%)
*Fisher Freeman Halton test or Fisher's exact test was used where cells counts < 5 (α=0.05)
† Chi-square tests were used (α=0.05)

Table 5. Weighted Inter-rater Kappa Statistic for Radiological Features

	All Readers	Readers A, B, C	Readers B, C, D	Readers A,C,D	Readers A, B, D
Overall Aggressiveness Rating	0.43 (0.30-0.54)	0.38 (0.30-0.44)	0.42 (0.40-0.47)	0.42 (0.30-0.54)	0.44 (0.40-0.47)
Degree Enhancement	0.48 (0.41-0.56)	0.49 (0.42-0.55)	0.42 (0.42-0.43)	0.49 (0.41-0.56)	0.46 (0.41-0.55)
Contour	0.51 (0.38-0.62)	0.56 (0.45-0.62)	0.59 (0.53-0.61)	0.49 (0.38-0.62)	0.45 (0.38-0.53)
Venous Invasion	0.63 (0.55-0.75)	0.69 (0.62-0.75)	0.61 (0.60-0.62)	0.63 (0.55-0.75)	0.62 (0.55-0.71)
Regional Lymph Node Involvement	0.42 (0.25-0.66)	0.46 (0.30-0.66)	0.58 (0.43-0.66)	0.38 (0.25-0.48)	0.33 (0.25-0.43)
Perilesional Fat Stranding[a]	0.33 (0.21-0.40)	0.21[b]	0.40[b]	0.37[b]	0.33 (0.21-0.40)
Tumor Necrosis	0.61 (0.54-0.69)	0.58 (0.54-0.67)	0.66 (0.64-0.67)	0.60 (0.54-0.69)	0.58 (0.54-0.64)
Tumor Consistency	0.32 (0.08-0.58)	0.42 (0.28-0.58)	0.30 (0.08-0.40)	0.38 (0.28-0.58)	0.21 (0.08-0.28)
Location	0.51 (0.42-0.60)	0.46 (0.42-0.51)	0.47 (0.46-0.50)	0.57 (0.51-0.60)	0.50 (0.42-0.59)
Hilar status	0.58 (0.43-0.72)	0.54 (0.43-0.64)	0.63 (0.61-0.64)	0.60 (0.53-0.72)	0.52 (0.43-0.61)
Collateral Vascularity	0.59 (0.49-0.66)	0.60 (0.57-0.62)	0.58 (0.49-0.62)	0.62 (0.57-0.66)	0.54 (0.49-0.57)

All values are reported as: mean (range)

[a] For perilesional fat stranding, reader C did not utilize one of the score categories thus all pairs that included reader C were removed from the analysis to calculate kappa statistics

[b] Only one pair was available for calculating the kappa statistic therefore range is not available

[c] The average simple kappa statistic is reported for Tumor Consistency since it is a binary variable

Table 6. Radiologic Feature Sum Scores Spearman Rank Correlation Matrix

	Overall Aggressive-ness Rating	Degree Enhance-ment	Contour	Venous Invasion	Regional Lymph Node Involvement	Perilesional Fat Stranding	Tumor Necrosis	Tumor Consistency	Location	Hilar status	Collateral Vascularity
Overall Aggressive-ness Rating	1.0000	0.5887	0.7472	0.6336	0.5597	0.6734	0.7717	0.3216	0.4315	0.6250	0.7518
Degree Enhancement		1.0000	0.3068	0.3517	0.1797	0.3568	0.5931	0.2921	0.2877	0.2677	0.4025
Contour			1.0000	0.5889	0.5114	0.5943	0.4911	0.1875	0.5117	0.5705	0.6449
Venous Invasion				1.0000	0.5844	0.4876	0.5264	0.1406	0.3786	0.5317	0.6333
Regional Lymph Node Involvement					1.0000	0.5324	0.3893	0.0420	0.4123	0.4779	0.6201
Perilesional Fat Stranding						1.0000	0.6194	0.1841	0.3021	0.3244	0.6602
Tumor Necrosis							1.0000	0.2952	0.2173	0.3890	0.6580
Tumor Consistency								1.0000	0.1850	0.1638	0.1752
Location									1.0000	0.6787	0.3535
Hilar status										1.0000	0.5546
Collateral Vascularity											1.0000

Table 7. Logistic Regression Model Selection

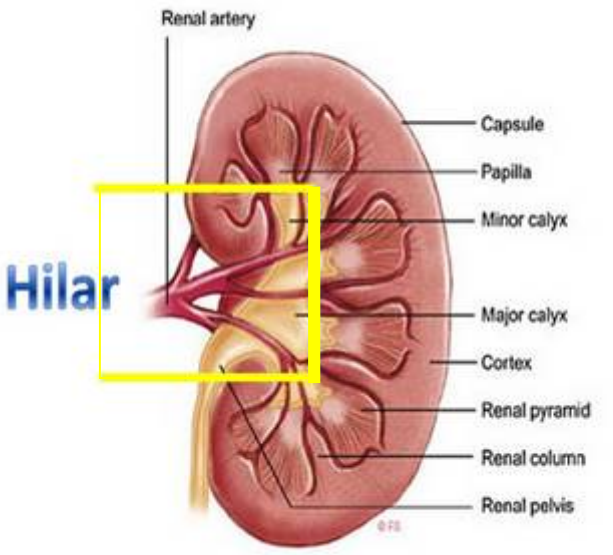
	Model 1			Model 2		
Covariate	**OR (95%CI)**		**p-value†**	**OR (95%CI)**		**p-value†**
OAR score sum	1.42	(1.13, 1.78)	0.0029†	1.44	(1.14, 1.82)	0.0021†
OAR range	1.01	(0.49, 2.08)	0.9851	1.07	(0.51, 2.23)	0.8668
Age (50-69 yrs. old)[a]	1.43	(0.58, 3.51)	0.4341	1.49	(0.60, 3.70)	0.3853
Age (70-89 yrs. old)[a]	1.38	(0.49, 3.87)	0.5442	1.52	(0.53, 4.33)	0.4381
Sex (Male)[a]	1.97	(0.95, 4.08)	0.0672	1.90	(0.89, 4.02)	0.0952
Tumor Size	1.12	(0.94, 1.33)	0.2192	1.10	(0.92, 1.32)	0.2802
Smoker (Current)[a]		--	--	0.79	(0.39, 1.61)	0.5188
Model Statistics						
Likelihood Ratio , χ^2 (p-value)	35.41 (<0.0001)			33.34 (<0.0001)		

† Test of significance used $\alpha = 0.05$, to predict high tumor grade (III/IV)

[a] Age referent = 30-49 yrs old, Sex referent = Female, Smoker referent = Not a current smoker

Figures

Figure 1. Hilar Box



In determining Hilar status, the readers were provided the image shown here. The hilar box is located within the highlighted area.