JOURNAL OF ENDOUROLOGY
Volume 31, Number 1, January 2017

Pp. 85–90
DOI: 10.1089/end.2016.0537

Endoscopic Evidence That Randall's Plaque is Associated with Surface Erosion of the Renal Papilla

Andrew J. Cohen, MD,[1] Michael S. Borofsky, MD,[2] Blake B. Anderson, MD,[1] Casey A. Dauw, MD,[2] Daniel L. Gillen, PhD,[3] Glenn S. Gerber, MD,[1] Elaine M. Worcester, MD,[4] Fredric L. Coe, MD,[4] and James E. Lingeman, MD, FACS[2]

Abstract

Objective: This study was conducted to assess the reliability and precision of an endoscopic grading scale to identify renal papillary abnormalities across a spectrum of equipment, locations, graders, and patients.
Materials and Methods: Intra- and interobserver reliability of the papillary grading system was assessed using weighted kappa scoring among 4 graders reviewing a single renal papilla from 50 separate patients on 2 occasions. Grading was then applied to a cohort of patients undergoing endoscopic stone removal procedures at two centers. Patient factors were compared with papillary scores on the level of the papilla, kidney, and patient.
Results: Graders achieved substantial (kappa >0.6) intra- and inter-rater reliability in scored domains of ductal plugging, surface pitting, and loss of contour. Agreement for Randall's Plaque (RP) was moderate. Papillary scoring was then performed for 76 patients (89 kidneys, 533 papillae). A significant association was discovered between pitting and RP that held both within and across institutions. A general linear model was then created to further assess this association and it was found that RP score was a highly significant independent correlate of pitting score ($F = 7.1$; $p < 0.001$). Mean pitting scores increased smoothly and progressively with increasing RP scores. Sums of the scored domains were then calculated as a reflection of gross papillary abnormality. When analyzed in this way, a history of stone recurrence and shockwave lithotripsy were strongly predictive of higher sums.
Conclusions: Renal papillary pathology can be reliably assessed between different providers using a newly described endoscopic grading scale. Application of this scale to stone-forming patients suggests that the degree of RP appreciated in the papilla is strongly associated with the presence of pitting. It also suggests that patients with a history of recurrent stones and lithotripsy have greater burdens of gross papillary disease.

Keywords: endoscopy, grading, papillae, pitting, Randall's plaque, ureteroscopy

Introduction

The renal papilla is critically important to stone pathogenesis and is the origin of many, if not most, calcium-based stones.[1] It is known to have variable appearances based upon different pathophysiologies and systemic stone-forming diseases.[2–5] The introduction and widespread utilization of high-definition digital ureteroscopes have allowed urologists an unparalleled ability to visualize the renal papilla at the time of endoscopic stone surgery.[6] Nonetheless, the significance of variable papillary appearance remains understudied and poorly understood.

Recently, two endoscopic papillary grading scales have been described with the intention of standardizing the description of papillary abnormalities.[7,8] The goal is to provide a reliable classification system that could ultimately be used to gain greater appreciation into how stones form as well as link papillary morphology to meaningful clinical endpoints. However, these systems have yet to be applied widely and to date only minimal single institution data have been provided. Furthermore, the mechanisms by which the abnormal papillary features occur and how they relate to each other remain unclear. We sought to assess both the reliability and clinical relevance of papillary scoring by formally analyzing consistency of scoring between urologists at two medical centers and then applying it to patients undergoing endoscopic stone surgery at each site.

Materials and Methods

Patients at two medical centers had endoscopic mapping and recording of their renal papillary anatomy at the time of surgery. The study was approved by the local Institutional

[1]Section of Urology, Department of Surgery University of Chicago, Chicago, Illinois.
[2]Department of Urology, Indiana University School of Medicine, Indianapolis, Indiana.
[3]Department of Statistics, Program in Public Health, and Department of Epidemiology, University of California, California, Irvine.
[4]Section of Nephrology, University of Chicago, Chicago, Illinois.

Review Board at each center (IRB 14-1111 and IRB 1010002261). Clinical characteristics and metabolic data, when available, were collected to correlate with endoscopic findings. Stones removed from patients were analyzed using either Micro CT or photomicroscopy and infrared spectroscopy (Beck Labs, Indianapolis, IN). Majority stone type was defined as 50% or more of total stone composition: calcium oxalate (CaOx), calcium phosphate (CaP), or uric acid (UA). Brushite stones were considered CaP for the purposes of the analysis. No cystine or struvite stones were included. Patient characteristics were compared using Fisher's exact and two-sample *t*-tests. All statistics were performed using Stata 13 (Statacorp, College Station, TX) and Systat 13.1 (San Jose, CA) with $p < 0.05$ considered significant.

We have recently proposed a papillary endoscopic grading system that has been described in detail.[7] Briefly, the grading scale measures papillary appearance in the domains of ductal plugging, surface pitting, loss of contour, and Randall's Plaque (RP). We have modified the scale to allow for statistical testing between all domains; as such, RP is now assigned a numeric value of 0, 1, or 2 in line with the other domains (Table 1).

Validation

Four urologists at two centers met for 90 minutes and reviewed example cases to ensure similar comprehension of the grading system. One senior level and one junior level urologist graded at each center. A validation study was designed to assess intra- and interobserver reliability between graders. Graders independently reviewed videos of a single renal papilla from 50 patients on 2 occasions. Each video lasted between 15 and 30 seconds. The validation cohort included videos made with both a Flex-X^c scope (KARL STORZ Endoscopy-America, Inc.) and ACMI DUR-D digital scope (Olympus Surgical Technologies America) at center 1 and center 2, respectively.

For this validity cohort, weighted kappa scores were calculated to assess intra- and interobserver reliability.[9] Interpretations of kappa scores are as follows: <0 less than chance agreement, 0.01 to 0.20 slight agreement, 0.21 to 0.40 fair agreement, 0.41 to 0.60 moderate agreement, 0.61 to 0.80 substantial agreement, and 0.81 to 0.99 almost perfect agreement.[10] The videos used for this validity exercise were not used for subsequent prospective analysis. Additionally, no patient data were studied from whom these videos were taken.

Analysis of interaction of grading scale domains

We then applied papillary grading to patients at the time of ureteroscopy. At center 1, 13 patients underwent bilateral ureteroscopy with complete papillary mapping and 63 patients at center 2 underwent unilateral ureteroscopy. A complete video was created using a Flex-X^c scope (KARL STORZ Endoscopy-America, Inc.) at center 1 and an ACMI digital scope (Olympus Surgical Technologies America) at center 2. Short clips of the papillae were made such that each papilla was featured. Two reviewers at each center applied the scoring system to each video. This was done in sequence for each patient at center 1 and in a completely randomized order at center 2.

After papillary scoring was completed, we sought to determine the associations between domains of ductal plugging, surface pitting, loss of contour, and RP. Spearman correlation coefficients were calculated for this purpose. Kruskal–Wallis testing was applied to compare subgroup median scores. Furthermore, we utilized backward-stepping general linear modeling (GLM) in a data-driven mode with Akaike information criterion to assess significant predictors of domain scores. This analysis pooled scores from all four graders; therefore, particular graders and institutions did not enter the model. Significance of adjusted means was assessed with Tukey's honestly significant difference test.

Analysis of papillary disease sum

Scores of each domain were added to create a final sum to test their ability to reflect gross papillary abnormality. By definition, the lowest possible score is 0 and maximum possible score is 8. Scores were calculated for each papilla. Papillae were then averaged across kidneys as well as across patients. We assessed clinical factors associated with higher sums.

Results

Validation

From the validation cohort of 50 videos, graders consistently achieved substantial intrarater agreement across all measured domains (kappa of 0.61–0.80) (Fig. 1). Likewise, inter-rater reliability was substantial in domains of plugging, pitting, and loss of contour. The amount of RP achieved moderate agreement.

Analysis across score domains

In the prospective cohort, 76 patients were enrolled (89 kidneys, 533 papillae). Patient characteristics are described in Table 2. Notably, all patients in center 1 were recurrent stone formers, were more likely to have undergone shockwave lithotripsy (SWL), and were more likely Caucasians. Analysis using Spearman correlations revealed that the

Table 1. Scale for Abnormal Papillary Appearance

Score	*0*	*1*	*2*
Plugging	0 Yellow plug deposits/ dilated ducts	≤5 Yellow plug deposits/ dilated ducts	>5 Yellow plug deposits/ dilated ducts
Pitting	None	≤25% Papillary surface involved	≥25% Papillary surface involved
Loss of contour	None	Depressed	Completely flattened
Amount of Randall's Plaque	Mild	Moderate	Severe
Final score		Sum	

Adopted with permission from Borofsky et al. (2015).

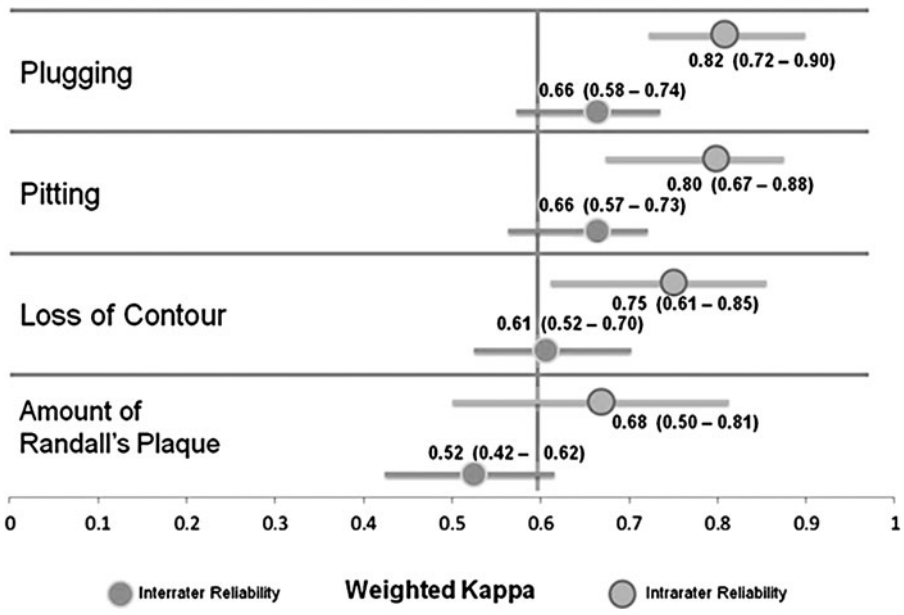

FIG. 1. Reliability among 4 graders for 50 papillary videos.

strongest interaction between domains was RP and pitting. Moderate correlation between pitting and RP was maintained on the level of each papilla (0.515), by patient (0.606), or averaged by kidney (0.606) (all $p < 0.05$). When analyzing data from each center separately, these relationships held. When removing patients with enteric disease, unknown stone type, or parathyroid disease, this relationship also persisted. An illustration of severe pitting and RP can be seen in Figure 2.

Other notable associations included a negative association of plugging and RP (–0.190) and a small positive association of contour loss with both pitting (0.283) and RP (0.149) only on the papillary unit level (all $p < 0.05$). All other relationships were independent of one another ($p > 0.05$). When analyzed by stone type, the relationships held for calcium stones, but not for UA stone formers. In general, center 1 detected higher overall amounts of RP with an average score of 1.2/2 for their patients compared with 0.4/2 for center 2 ($p < 0.001$). In contrast, center 2 noted more plugging, with an average score of 0.7/2 *vs* 0.3/2 across all papillae ($p < 0.001$).

Given the consistent correlation seen between pitting and RP, we explored the relationship further with parametric analysis. In each center and within each reader, increasing RP scores were associated with increasing pitting scores. (Fig. 3) In a GLM with the pooled pitting score as dependent variable and age, race, SWL, stone recurrence, stone composition, plugging score, and RP as categorical variables, we found

Table 2. Patient Characteristics by Center

	Center 1 (%) N = *13* *Patients*	*Center 2 (%)* N = *63* *Patients*	p
Papillae	217	316	—
Kidney	26	63	—
Gender (male)	69.2%	52.4%	0.211
Age (median, IQR)	46 (42–52)	51 (41–63)	0.14
Race (%)			
Caucasian	100	47.6	0.008
African American	0	39.7	
Asian	0	3.1	
Other	0	9.5	
BMI (median, IQR)	30.3 (24.7–33.7)	28.9 (24.5–37.3)	0.93
Majority stone type (%)			
CaOx	83.3	52.4	0.325
CaP	16.7	31.8	
UA	0	15.9	
Etiology (%)			
Parathyroid disease	7.7	9.5	1.00
Enteric disease	0	19.1	0.133
Recurrent stone	100	54.2	0.001
History of SWL	38.5	15.9	0.075

CaOx = calcium oxalate; CaP = calcium phosphate; SWL = shockwave lithotripsy; UA = uric acid.

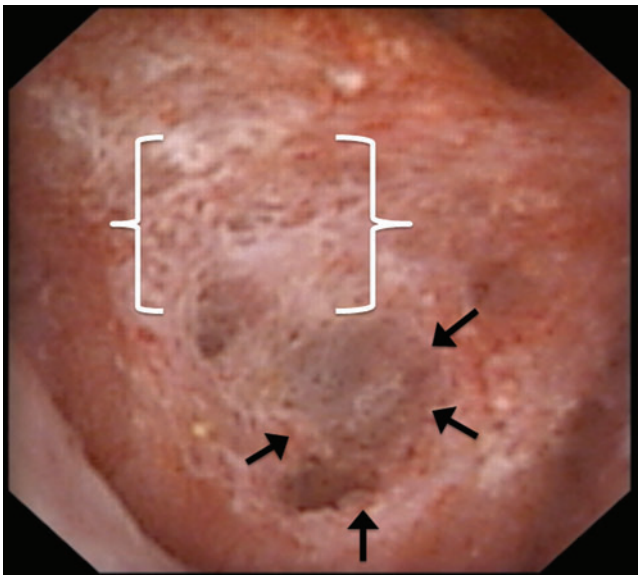

FIG. 2. Illustrative still of papilla demonstrating concurrent RP (white brackets) and pitting (area highlighted by black arrows).

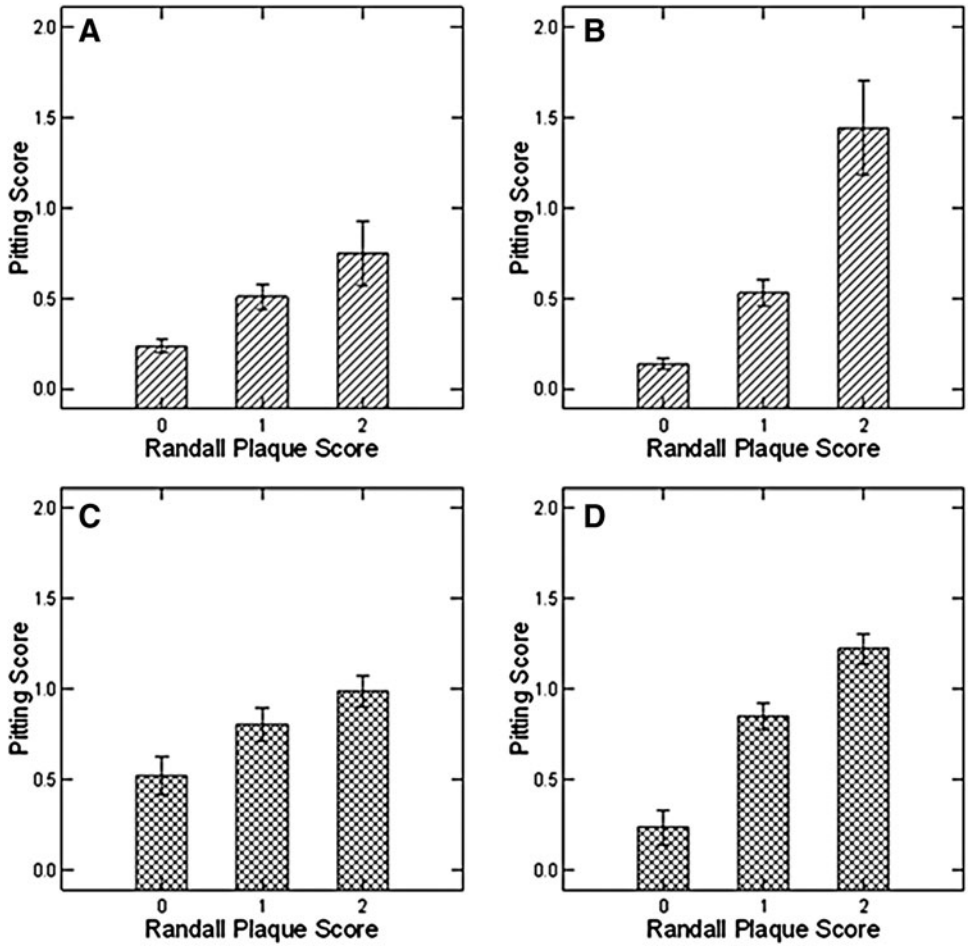

FIG. 3. Mean pitting scores by plaque. **(A)** Grader 1 Center 2, **(B)** Grader 2 Center 2, **(C)** Grader 1 Center 1, **(D)** Grader 2 Center 1. On each graph, differences between bars were significant on ANOVA except for the following: *A*, RP score of 2 does not differ from a score of 1, and *C*, RP score of 0 and 2 did not differ from 1, but they differ from each other.

that RP score was a highly significant independent correlate of pitting ($F = 7.1$; $p < 0.001$). Fully adjusted for recurrence and SWL, mean pitting scores increase smoothly and progressively with increasing RP scores (Fig. 4). The remaining variables were not significant predictors of pitting score and did not require adjustment (Table 3).

Analysis of score totals

Total scores, a potential reflection of gross papillary abnormality, were significantly different at each site with a median of 3/8 (IQR: 2–4) at center 1 *vs* 1.5/8 (IQR: 0.5–2.5) at center 2 ($p < 0.001$). Total scores were higher for those patients with a history of prior stones [2.5 (IQR 1.5–4) *vs* 1(IQR: 0.5–2); $p < 0.001$]. SWL was collinear with recurrence; with that in mind, a history of SWL was associated with increased total scores. [3 IQR:(2–4) *vs* 1.5 IQR:(1–3); $p < 0.001$]. There were fewer recurrent stone formers at center 2 (54% vs. 100%), but at center 2, scores for recurrent stone formers were still consistently higher [1.75 (IQR: 1–3) *vs* 1 (IQR:0.5–2); $p = 0.01$].

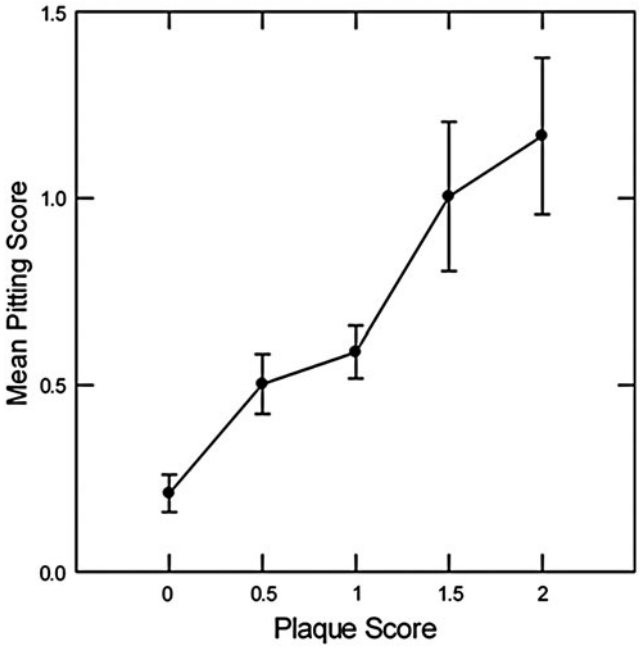

FIG. 4. Linear relationship between mean pitting and plaque score fully adjusted for recurrence and SWL, trend significant <0.001. Together, recurrence and SWL were powerfully associated with pitting ($F = 18$; $p < 0.001$). SWL, shockwave lithotripsy.

Discussion

We sought to investigate the reliability of a recently described papillary grading system.[7] Our findings demonstrate that the system can be used to reliably quantify the magnitude of papillary abnormalities. Additionally, application of this system provides unique insight into potential mechanisms of

Table 3. Factors Associated with Pitting Corrected for History of Recurrence and SWL

Variable[a]	*F- ratio*	p
Degree of RP	24.0	<0.01
History of SWL	20.4	<0.01
Interaction of history of recurrence and SWL	18.0	<0.01
Interaction of history of recurrence and degree of RP	3.93	<0.01
Interaction of history of SWL and degree of RP	3.32	0.01

[a]Stepwise selection eliminated contour score, stone type, plugging score, recurrence, and their interaction variables due to $F < 1.0$ and $p > 0.2$.
RP = Randall's Plaque.

stone formation as well as associations between clinical factors and papillary pathology previously not explored on a large scale. This methodology is particularly timely as the use of ureteroscopy as a primary treatment of upper urinary tract stones continues to grow at a rapid rate.[11] Specifically, establishing reference standards for normal and abnormal papillary appearance could help substratify stone formers.

We currently lack reliable tools to risk stratify stone formers. The ROKS nomogram has shown promise, but is based mainly on data related to an acute stone episode that essentially provide a snapshot in time.[12] Papillary appearance presumably provides details related to the duration and severity of disease, as indicated by the degree of visible pathology. Other methods of classifying stone formers such as 24-hour urine collections, the mainstay of metabolic evaluation, are variable based on patient compliance and diet. Likewise, stone composition determination by traditional techniques such as photomicroscopy and Fourier transform infrared spectroscopy is subject to error.[13,14] In contrast, papillary pathology may be a more stable and reliable indicator of disease severity as it is potentially static and objective if measured using a validated instrument. Patients observed intraoperatively to have a high burden of papillary pathology could be considered for more aggressive surveillance or further metabolic evaluation. The fact that recurrence and SWL history were associated with higher scores is an early suggestion of the utility of our scale to risk stratify patients. At least for SWL, animal models suggest that papillary injury induces cellular fragmentation and necrosis and it would not be surprising for multiple treatments to manifest in more detectable damage on endoscopy.[15]

Similarly, it is likely that certain types of stone formers have corresponding patterns of papillary abnormalities. For example, the relationship between pitting and RP was exclusive to calcium stone formers, but did not hold for UA stone formers, indicating separate mechanisms of stone formation. This mirrors the current hypothesis that UA precipitates in solution rather than being affixed to the papillary surface.[16] As the concept of papillary grading continues to evolve, it is quite possible that endoscopic papillary phenotypes might have a supplementary role to standard techniques of classifying stone formers in guiding more precise preventative strategies and treatments.

Further recognition of papillary pathology could also set the stage for new theories of stone formation. Our finding that pitting is associated with higher degrees of RP is evidence for this concept as such observations would be impossible without a descriptive terminology. For example, how are pits formed, and what is their relationship to RP? One possibility is that pits are formed when RP is pulled away by passage or removal of a stone. Does urothelium regrow over the exposed surface? Is an exposed surface equally susceptible to RP formation? Is the pathophysiology of stone formation expedited or slowed as a result of the erosion? All of these potentially important questions can only be addressed if we provide the language to make such characterizations possible.

Our work is the largest formal assessment of papillary pathology to date. Moreover, the study of relationships between features of papillary pathology measured in each domain is novel. Many of these features have been well described by Evan et al., but relationships between factors (i.e., plugging and RP) have only been studied on a smaller scale among highly selected patient populations.[2–5] It has been previously noted that non-stone formers have no visual papillary abnormalities.[1,8] Ultimately, we found that pitting and RP are positively correlated, whereas plugging is negatively associated with RP. This lends support to the concept that stones may form through two independent (or divergent) mechanisms. Patients with CaP stones have previously been observed on biopsy to have collecting ducts filled with crystal deposits or plugging.[4] In contrast, patients with idiopathic CaOx stones have stone overgrowth on RP.[17] This has yet to be fully validated, but efforts such as these will be necessary to gain greater insight into whether this will hold true in a general population of stone formers.

Limitations

The interobserver agreement between variables in the grading scale was not perfect. Given that static epidemiologic data extracted from patient charts often only attain moderate agreement among graders, we believe our preliminary reliability to be promising.[18,19] Furthermore, our findings reflect the initial experience using this scale. Such efforts at creating and applying grading scales are likely to improve with experience; for example, precision and reliability of the Gleason scoring system, used routinely by pathologists to characterize the aggressive potential of prostate cancer, have been shown to improve over time with training and exposure.[20] Kappa for a given reading of Gleason scores by trained pathologists was 0.67 in a recent study.[21] This demonstrates that a scoring system need only be able to demonstrate substantial agreement to be clinically relevant. There were significant differences between the degree of pathology seen at both institutions, with patients from center 1 having higher degrees of RP and center 2 having higher degrees of plugging. We suspect that this may be due to different patient populations as center 1 specifically recruited patients undergoing bilateral procedures and thus more likely to have a greater degree of stone burden and more aggressive disease. However, this can also be seen as a strength given the fact that the relationships between pitting and RP were maintained across each center regardless of baseline disease. Finally, we acknowledge that this grading scale requires further study among larger numbers of patients and validation among the wider urologic community. However, initial efforts such as the one described will be necessary in appropriately refining the scale to ensure it is correctly capturing the intended information and being interpreted with the greatest degree of accuracy.

Conclusion

We provide evidence of the reliability and validity of a grading system for renal papillary damage. We also note that a relationship between RP and pitting holds despite heterogeneity in patients, graders, and equipment. Pitting, identified on endoscopy, appears to be independent of plugging in any given papillae, suggesting that each of these are unique manifestations of papillary pathology and likely have entirely separate associations with stone pathogenesis. Total scores, a potential marker of overall papillary health, are correlated with recurrence and SWL, suggesting that as the degree of stone disease gets worse, abnormalities of the papillae become more common. This grading system may have clinical utility as a tool for research as well as intraoperative risk stratification. As we begin to understand the implication of papillary abnormalities, we are optimistic that it may be used as a surrogate mechanism to tailor specific treatment strategies to afflicted patients.

Acknowledgment

This work was supported by NIH P01 DK-56788.

Author Disclosure Statement

Dr. Lingeman is a consultant/advisor, investor, meeting participant/lecturer, and scientific study trial participant for Boston Scientific Corp., and owner, medical director, for Beck Analytical; Dr. Coe is a consultant for Labcorp; and for the remaining authors, no competing financial interests exist.

References

1. Evan AP, Lingeman JE, Coe FL, Parks JH, Bledsoe SB, Shao Y, et al. Randall's plaque of patients with nephrolithiasis begins in basement membranes of thin loops of Henle. J Clin Invest 2003;111:607–616.
2. Evan AE, Lingeman JE, Coe FL, Miller NL, Bledsoe SB, Sommer AJ, et al. Histopathology and surgical anatomy of patients with primary hyperparathyroidism and calcium phosphate stones. Kidney Int 2008;74:223–229.
3. Evan AP, Lingeman JE, Worcester EM, Bledsoe SB, Sommer AJ, Williams JC, et al. Renal histopathology and crystal deposits in patients with small bowel resection and calcium oxalate stone disease. Kidney Int 2010;78:310–317.
4. Evan AP, Lingeman JE, Coe FL, Shao Y, Parks JH, Bledsoe SB, et al. Crystal-associated nephropathy in patients with brushite nephrolithiasis. Kidney Int 2005;67:576–591.
5. Evan AP, Lingeman J, Coe F, Shao Y, Miller N, Matlaga B, et al. Renal histopathology of stone-forming patients with distal renal tubular acidosis. Kidney Int 2007;71:795–801.
6. Humphreys MR, Miller NL, Williams JC, Evan AP, Munch LC, Lingeman JE. A new world revealed: Early experience with digital ureteroscopy. J Urol 2008;179:970–975.
7. Borofsky MS, Paonessa JE, Evan AP, Williams JC, Coe FL, Worcester EM, et al. A proposed grading system to standardize the description of renal papillary appearance at the time of endoscopy in patients with nephrolithiasis. J Endourol Endourol Soc 2015. doi: 10.1089/end.2015.0298.
8. Almeras C, Daudon M, Ploussard G, Gautier JR, Traxer O, Meria P. Endoscopic description of renal papillary abnormalities in stone disease by flexible ureteroscopy: A proposed classification of severity and type. World J Urol 2016. doi: 10.1007/s00345-016-1814-6.
9. Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968;70:213–220.
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–174.
11. Oberlin DT, Flum AS, Bachrach L, Matulewicz RS, Flury SC. Contemporary surgical trends in the management of upper tract calculi. J Urol 2015;193:880–884.
12. Rule AD, Lieske JC, Li X, Melton LJ, Krambeck AE, Bergstralh EJ. The ROKS nomogram for predicting a second symptomatic stone episode. J Am Soc Nephrol JASN 2014;25:2878–2886.
13. Daudon M, Donsimoni R, Hennequin C, Fellahi S, Le Moel G, Paris M, et al. Sex- and age-related composition of 10 617 calculi analyzed by infrared spectroscopy. Urol Res 1995;23:319–326.
14. Krambeck AE, Lingeman JE, McAteer JA, Williams JC. Analysis of mixed stones is prone to error: A study with US laboratories using micro CT for verification of sample content. Urol Res 2010;38:469–475.
15. Shao Y, Connors BA, Evan AP, Willis LR, Lifshitz DA, Lingeman JE. Morphological changes induced in the pig kidney by extracorporeal shock wave lithotripsy: Nephron injury. Anat Rec A Discov Mol Cell Evol Biol 2003;275:979–989.
16. Sakhaee K. Epidemiology and clinical pathophysiology of uric acid kidney stones. J Nephrol 2014;27:241–245.
17. Evan AP, Coe FL, Lingeman JE, Shao Y, Sommer AJ, Bledsoe SB, et al. Mechanism of formation of human calcium oxalate renal stones on Randall's plaque. Anat Rec Hoboken NJ 2007;290:1315–1323.
18. Horwitz RI, Yu EC. Assessing the reliability of epidemiologic data obtained from medical records. J Chronic Dis 1984;37:825–831.
19. Eder C, Fullerton J, Benroth R, Lindsay SP. Pragmatic strategies that enhance the reliability of data abstracted from medical records. Appl Nurs Res ANR 2005;18:50–54.
20. Nakai Y, Tanaka N, Shimada K, Konishi N, Miyake M, Anai S, et al. Review by urological pathologists improves the accuracy of Gleason grading by general pathologists. BMC Urol 2015;15:70.
21. Sadimin ET, Khani F, Diolombi M, Meliti A, Epstein JI. Interobserver reproducibility of percent gleason pattern 4 in prostatic adenocarcinoma on prostate biopsies. Am J Surg Pathol 2016. doi: 10.1097/PAS.0000000000000714.

Address correspondence to:
James E. Lingeman, MD, FACS
Department of Urology
Indiana University School of Medicine
1801 North Senate Blvd., Suite 220
Indianapolis, IN 46202

E-mail: jlingeman@iuhealth.org

Abbreviations Used

CaOx = calcium oxalate
CaP = calcium phosphate
CT = computed tomography
GLM = general linear modeling
RP = Randall's Plaque
SWL = shockwave lithotripsy
UA = uric acid