Evaluation and Comparison of Urolithiasis Scoring Systems in Percutaneous Kidney Stone Surgery

Kevin Labadie,* Zhamshid Okhunov,* Arash Akhavein, Daniel Moreira, Jorge Moreno-Palacios, Michael del Junco, Zeph Okeke, Vincent Bird, Arthur D. Smith and Jaime Landman†

From the Departments of Urology, University of California-Irvine (KL, ZO, JMP, MdJ, JL), Orange, California, and University of Florida (AA, VB), Gainesville, Gainesville, Florida, and Smith Institute for Urology, North Shore LIJ Health System (DM, ZO, ADS), New Hyde Park, New York

Purpose: Contemporary predictive tools for percutaneous nephrolithotomy outcomes include the Guy stone score, S.T.O.N.E. nephrolithometry and the CROES nephrolithometric nomogram. We compared each scoring system in the same cohort to determine which was most predictive of surgical outcomes.

Methods: We retrospectively reviewed the records of patients who underwent percutaneous nephrolithotomy between 2009 and 2012 at a total of 3 academic institutions. We calculated the Guy stone score, the S.T.O.N.E. nephrolithometry score and the CROES nephrolithometric nomogram score based on preoperative computerized tomography images. A single observer at each institution reviewed all images and assigned scores. Univariate and multivariate analysis was done to determine the most predictive scoring system.

Results: We enrolled 246 patients in study. In stone-free patients vs those with residual stones the mean Guy score was 2.2 vs 2.7, the mean S.T.O.N.E. score was 8.3 vs 9.5 and the mean CROES nomogram score was 222 vs 187 (each $p < 0.001$). Logistic regression revealed that the Guy, S.T.O.N.E. nephrolithometry and CROES nomogram scores were significantly associated with stone-free status ($p = 0.02$, 0.004 and <0.001, respectively). The Guy and S.T.O.N.E. nephrolithometry scores were associated with estimated blood loss ($p < 0.0001$ and 0.03) and length of stay ($p = 0.03$ and 0.009, respectively). The CROES nomogram did not predict estimated blood loss or length of stay.

Conclusions: All scoring systems and the stone burden equally predicted stone-free status. The Guy and S.T.O.N.E. nephrolithometry scores were associated with estimated blood loss and length of stay. A single scoring system should be adopted to unify reporting.

Key Words: kidney; urolithiasis; nephrostomy, percutaneous; nomograms; research design

Abbreviations and Acronyms

CROES = Clinical Research Office of Endourological Society

CT = computerized tomography

EBL = estimated blood loss

LOS = length of stay

OT = operative time

PCNL = percutaneous nephrolithotomy

SFS = stone-free status

S.T.O.N.E. = stone size, tract length, obstruction, number of involved calyces and essence or stone density

Accepted for publication July 22, 2014.

Study received institutional review board approval.

* Equal study contribution.

† Correspondence: Department of Urology, University of California-Irvine, 333 City Blvd. West, Suite 2100, Orange, California 92868 (telephone: 714-456-3330; e-mail: landmanj@uci.edu).

THERE has been a marked increase in the prevalence of kidney stone disease in the United States in the last 2 decades, approaching 7% in females and 10.3% in males in 2010.[1] With this dramatic increase in stone disease incidence and prevalence the use of PCNL to treat a large stone burden has continued to increase.[2–4] Despite continuous refinements in surgical techniques and technology the overall complication rate of PCNL has also increased.[5] An accurate estimate of treatment success is crucial

0022-5347/15/1931-0001/0
THE JOURNAL OF UROLOGY®

http://dx.doi.org/10.1016/j.juro.2014.07.104
Vol. 193, 1-6, January 2015
Printed in U.S.A.

for optimal decision making and informed patient counseling.

To characterize kidney stone complexity preoperative radiological evaluation with CT has become common practice in the United States. CT provides high resolution spatial imaging for accurate characterization of the stone size and distribution, pelvicalyceal anatomy, anomalies and anatomical relationships that may dictate the feasibility and risks of different treatment modalities.

With these measurable stone and patient features the Guy stone score,[6] S.T.O.N.E. nephrolithometry[7] and the CROES nephrolithometry nomogram[8] were introduced for systematic and quantitative assessment of kidney stones. In addition to imaging characteristics, these models also take into account other patient features that contribute to disease outcome, such as obesity, renal surgical history, spinal cord injury and spina bifida status, as well as surgeon experience. These parameters are thought to provide the surgeon with an assessment of the complexity and intricacy of each patient. The scoring systems serve as disease stratification tools that allow the surgeon to more accurately predict PCNL outcomes to improve patient counseling and surgical planning.[6–8]

Another potential advantage of scoring systems is uniform and standardized reporting across different series. To date comparative evaluation of treatment for urolithiasis has been limited by the lack of a widely accepted standardization system.[9,10] Uniform academic and clinical reporting would empower physicians to better compare data from different institutions and improve the overall quality of urological research.

To date there has been no direct comparison of the existing scoring systems.[6–8] Comparison and analyses of these tools support refinements and improvements in these systems, which may ultimately facilitate the creation of a more universal and widely accepted scoring system. Thus, we evaluated and compared these scoring systems to assess their relative predictive value for surgical outcomes. We also reviewed the features of each system, similarities and differences, applicability in clinical practice and relevance in academic reporting.

METHODS

After obtaining institutional board review approval we retrospectively reviewed the charts of patients treated with PCNL between 2009 and 2012 at a total of 3 academic institutions.

Selection Criteria

Study exclusion criteria included age less than 18 years, a history of surgery on the ipsilateral kidney, nephrostomy tube or stent placement in the ipsilateral kidney preoperatively and no available preoperative CT images. Patients who underwent repeat PCNL for recurrent stones on the ipsilateral kidney were included in analysis. If a patient underwent bilateral procedures, we selected 1 side at random to improve the independence of data points.

Measurements

We calculated the Guy Score, S.T.O.N.E. nephrolithometry and the CROES nephrolithometric nomogram on all patients based on preoperative CT images, as described by Thomas,[6] Okhunov[7] and Smith[8] et al, respectively. A single observer from each institution reviewed all images and performed scoring according to each system. We provided standardized instructions to all reviewers on the application of each scoring system before data collection.

Perioperative Data

We collected patient demographic, clinical, perioperative and followup data in retrospective fashion. Collected information included age, gender, body mass index, surgical and medical history, renal anomalies, ASA (American Society of Anesthesiologists) score, EBL, fluoroscopy time, OT, stone location and size, number of renal punctures, number and location of dilated tracts, intraoperative and postoperative complications within 30 days, and LOS.

Outcomes

The primary study outcome was a comparison of the ability of the Guy score, S.T.O.N.E. nephrolithometry and the CROES nomogram to predict stone-free rates after PCNL. We defined stone-free status in our study as absent residual stones or stone fragments less than 2 mm at the termination of the procedure on as confirmed by postoperative CT.[11,12] CT was performed in all patients before discharge home or within 3 months postoperatively.

The secondary outcome was to evaluate the ability of the scoring systems to predict perioperative and postoperative complications within 30 days of the procedure. We classified all intraoperative and postoperative complications according to the modified Clavien system.[13] We also evaluated perioperative variables such as fluoroscopy time, OT, EBL and LOS.

Our surgical techniques were described previously.[14,15] Groups at all participating institutions had substantial experience with the PCNL procedure and the surgical technique was performed in similar fashion at the 3 academic institutions.

Statistical Analysis

We divided patients into 2 groups based on postoperative SFS. Baseline characteristics were compared between stone-free and nonstone-free patients using the chi-square test for categorical variables and the Student t-test for continuous data. The Guy score and the CROES nomogram were used in 4 groups each and S.T.O.N.E. nephrolithometry was used in 3. Descriptive statistics were used to show the stone-free rate across the 4 groups for each scoring system. ROC curves were generated for each scoring system and for the stone burden, which was measured in mm^2. The AUC and asymptotic 95% CI were calculated for each ROC curve. All statistical analysis was

2-tailed and done with Stata® 12.0 and R 3.0.1 (http://www.r-project.org/) with p <0.05 considered statistically significant.

RESULTS

We identified 246 patients who underwent PCNL between 2009 and 2012 and met study inclusion criteria. Table 1 lists patient demographics and stone characteristics.

Perioperative Data

The overall single procedure stone-free rate in the study was 56%. Mean stone size in stone-free patients and patients with residual stones was 654 and 1,525 mm^2, respectively (p <0.001). Overall 42 patients (17%) experienced postoperative complications, including Clavien grade I (fever or pain management with nonsteroidal anti-inflammatory drugs) in 23, Clavien grade II (fever treated with antibiotics or acute kidney injury managed with intravenous fluids) in 8, Clavien grade IIIA (obstruction requiring nephrostomy tube placement or Double-J® stent causing infundibular rupture) in 5, Clavien grade IIIB (significant bleeding requiring angioembolization, bleeding requiring nephrectomy or renal abscess treated with nephrectomy in the postoperative period) in 3 and Clavien grade IVA (acute kidney injury hemodialysis or septic shock) in 3. There were no deaths.

Table 1. *Patient demographic and clinical characteristics*

	Stone Free		Not Stone Free		p Value
No. pts (%)	137	(56)	107	(44)	—
Mean ± SD age	55.7 ±	14.9	55.1	(15.0)	0.774
No. male (%)	68	(63)	40	(37)	0.056
No. female (%)	69	(51)	67	(49)	
No. ASA score (%):					
1	13	(68)	6	(32)	0.432
2	46	(52)	42	(48)	
3	60	(59)	42	(41)	
4	4	(80)	1	(20)	
No. laterality (%):					
Lt	77	(58)	55	(42)	0.455
Rt	60	(54)	52	(46)	
Mean ± SD body mass index (kg/m^2)	30.6 ±	8.6	30.2 ±	7.7	0.711
No. punctures (%)	1.0	(0.1)	1.2	(0.6)	0.009
Mean ± SD operative time (mins)	122 ±	53	152 ±	65	<0.001
Mean ± SD EBL (ml)	68 ±	71	101 ±	113	0.005
Mean ± SD LOS (days)	3.1 ±	2.7	3.2 ±	2.7	0.608
Mean ± SD stone size (mm^2)	654	± 617	1,525	± 1,869	<0.001
No. calyces (%)	1.86	(1—4)	2.5	(1—5)	0.001
No. staghorn stones (%)	19		36		0.001
Mean ± SD HU essence (range)	937	(305—1,580)	937	(389—1,849)	0.435
No. location (%):					
Upper pole	41	(53)	36	(47)	0.135
Mid pole	24	(59)	17	(41)	
Lower pole	68	(61)	44	(39)	
Multiple	4	(29)	10	(71)	

Scoring Systems

In patients who were stone free and those with residual stones the mean Guy score was 2.2 and 2.7, the mean S.T.O.N.E. nephrolithometry score was 8.3 and 9.5 and the mean CROES nomogram score was 222 and 187, respectively (each p <0.001). Table 2 shows the stone-free rate of each scoring system. The Guy score, S.T.O.N.E. nephrolithometry and CROES nomogram groups were significantly associated with SFS (p = 0.002, 0.004 and <0.001, respectively).

Table 3 and the figure show AUC and ROC curves for each scoring system and the stone burden. All scoring systems had similar accuracy and none was more predictive of SFS than the stone burden alone. The Guy score and S.T.O.N.E nephrolithometry were significantly associated with EBL (p <0.0001 and 0.03) and LOS (p = 0.03 and 0.009, respectively) but the CROES nomogram was not significantly associated with complications, EBL or LOS.

DISCUSSION

Tools that aid in prediction and decision making are different in design and methodology. Nomograms, risk groups, probability tables, and classification and regression tree analyses are the few most commonly applied examples. Although the 3 scoring systems evaluated in this study have different developmental concepts, they are meant to predict stone-free rates and complications while serving as disease stratification tools that provide the surgeon and patient with information on individual procedure complexity.[6–8]

Cumulatively the Guy score, S.T.O.N.E. nephrolithometry and the CROES nomogram incorporate 11 variables, of which only 4 are shared, including stone location, size and number, and staghorn status. The other 7 variables (tract length, renal pelvic obstruction, stone density, case volume

Table 2. *Stone-free rate of 3 scoring systems*

Scoring System	No. Stone Free/Total No. (%)	
Guy (grade):		
1	33/47	(70.2)
2	53/81	(65.4)
3	37/77	(48.1)
4	14/39	(35.9)
S.T.O.N.E. (category):		
5—6	24/34	(70.6)
7—8	53/80	(66.3)
9—13	70/130	(46.2)
CROES nomogram:		
80—129	5/22	(22.7)
130—169	26/56	(46.4)
170—219	25/55	(45.5)
220 or Greater	80/110	(72.7)

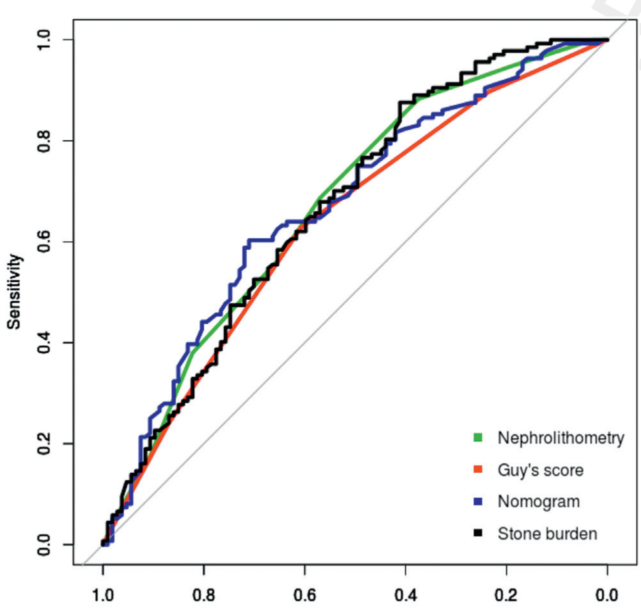
Table 3. *ROC curve values by scoring system*

Scoring System	ROC Curve	Roc
Guy score	0.634	0.566—0.702
S.T.O.N.E. nephrolithometry	0.670	0.602—0.738
CROES nomogram	0.671	0.602—0.739
Stone burden	0.668	0.599—0.737

per year, number of stones, treatment history, renal anatomy and spina bifida or spinal injury) are included separately in each scoring system. Additional key differences are the method by which each accounts for patient anatomical features. The Guy score includes abnormal renal anatomy and calyceal diverticulum. S.T.O.N.E. nephrolithometry and the CROES nomogram do not consider renal anomalies but S.T.O.N.E. nephrolithometry accounts for the other anatomical features mentioned. However, results of large-scale studies demonstrated that abnormal renal anatomy is not associated with inferior surgical outcomes.[16,17]

We compared the 3 scoring systems in a single cohort of patients who underwent PCNL. Although the heterogeneity of the scoring systems make analysis and direct comparison complex, they can be compared generally in a single patient cohort using reliable statistical methods. To our knowledge our study provides the first comparison of the 3 scoring systems in the same patient cohort.

We noted equal capacity of the Guy score, S.T.O.N.E. nephrolithometry and the CROES nephrolithometric nomogram to predict SFS. In



ROC curve of S.T.O.N.E. nephrolithometry, Guy score, CROES nomogram score and stone size.

current series none of the scoring systems was more predictive of SFS than the stone burden alone. This finding contradicts the initial studies of these classification systems.[6,7] The discrepancy may be a function of the fact that each system was constructed based on the population of patients tested, which introduced an intrinsic biased favoring predictive efficacy. Consistent with all previous reports, stone size undoubtedly remains the leading predictor of perioperative outcomes.

In addition to the primary outcome, the Guy score and S.T.O.N.E. nephrolithometry were significantly associated with perioperative outcomes. Overall these scoring systems attempt to incorporate important variables in an efficient, simple manner to quantitate renal stone complexity. Given the similar predictive abilities of the 3 systems, it is up to urologists to decide which should be implemented and used in clinical practice and academic reporting.

The Guy score and S.T.O.N.E. nephrolithometry were externally validated in numerous published studies. External validation of the Guy score was reported in 2 separate series by Mandal[18] and Ingirmasson[19] et al, in which the scoring system effectively predicted SFS. S.T.O.N.E. nephrolithometry was externally validated in a multi-institutional study of 850 patients.[20] The model was significantly associated with SFS, the overall complication rate ($p = 0.008$), EBL ($p = 0.001$), OT ($p < 0.001$) and LOS ($p = 0.016$). Akhavein et al evaluated S.T.O.N.E. nephrolithometry in 117 patients using strict surgical outcome criteria.[21] The stone-free rate was 75% and the S.T.O.N.E. score ranged from 6 to 12. In a logistic regression model the scoring system was significantly associated with SFS. The Guy score and S.T.O.N.E. nephrolithometry have excellent interobserver reliability.[19,22] To our knowledge the CROES nomogram has yet to be externally validated to date.

The Guy score and S.T.O.N.E. nephrolithometry use risk groups to determine the risk of an event. The categories of potential risk groups allow for improved differential stratification and selection of homogeneous patients who serve as a benchmark to assess the quality of various interventions in the effort to achieve superior patient care and outcomes. Although grouping homogenous patients into risk groups enables discrimination of those at low, medium and high risk, this methodology is associated with the assumption that patients in a risk group are equal. The initial report of S.T.O.N.E. nephrolithometry demonstrated that each increase in score is associated with 1.5 times more likelihood of a complication. Patients with a S.T.O.N.E. score of 9 to 13, who represent a high risk group, are at different risks for adverse events. This differs from the Guy score, which shows significant overlap in the way that patients are graded. For example, patients with partial and complete staghorn calculi are classified into grades 3 and 4, respectively. Given the poor, vague definitions of partial and complete staghorn stones, significant overlap and variations may potentially under or over grade the case, thus, decreasing scoring system accuracy.

Thomas et al highlighted this fact in their original study.[6] Their data revealed poor interobserver agreement when reviewers graded patients with partial vs complete staghorn stones. In contrast, nomograms have shown superior performance in other areas of urological research.[23,24] However, the lack of validation data as well as the large continuous scale of the CROES nomogram makes it difficult and impractical to implement it in a busy clinical routine.

The imaging modalities with which the scoring systems were developed also show inconsistencies. Since preoperative CT is the gold standard imaging modality, it is important that these scoring systems may be used with CT and were validated based on CT images. The Guy score and the CROES nomogram were initially developed using abdominal x-ray. In contrast, S.T.O.N.E. nephrolithometry is based on CT and consists of variables that are obtained specifically from CT images, making it best suited for use with contemporary imaging modalities. Stone size is an example of a variable that is easily and most accurately measured by CT that is not taken into account by the Guy score. Other important variables such as tract length, stone density and hydronephrosis severity are also measured exclusively on CT and only incorporated into S.T.O.N.E. nephrolithometry.

When considering the optimal scoring system, it is essential that it must be reproducible, easily implemented and adequately comprehensive for thorough reporting and comparison. When considering all limitations, we believe that S.T.O.N.E. nephrolithometry provides more accurate risk stratification data than the Guy score and offers easier application than the CROES nomogram. After carefully reviewing these systems we believe that S.T.O.N.E. nephrolithometry is the most comprehensive scoring system while remaining simple to implement in daily practice. It is easily remembered and applied with a simple acronym that is reproducible.[22] Furthermore, S.T.O.N.E. nephrolithometry is the only scoring system developed strictly using CT, which is routinely performed in almost all patients with urolithiasis. While the Guy score is also easy to implement, it has only 4 grades and provides limited information on disease extent.[25] This hinders its ability to stratify disease complexity, thus, limiting its usefulness for academic reporting and patient education.[26]

Although limitations are inherent due to the retrospective design of this study, we minimized these limitations with standardized data collection methods, and strict outcomes definitions and followup protocols.[11,12] Another possible study limitation was our exclusion criteria, which included patients with prior ipsilateral surgery and those with a stent or nephrostomy tube placed before the procedure. These cases were excluded since they are often more complex and do not reflect the typical PCNL experience. In addition, our data represent the experience of fellowship trained surgeons from 3 academic centers. Although surgical techniques may differ slightly, our results support the generalizability of the data.

CONCLUSIONS

The Guy score, S.T.O.N.E. nephrolithometry and the CROES nomogram were equally predictive of SFS in patients undergoing PCNL. The Guy score and S.T.O.N.E. nephrolithometry were associated with EBL and LOS. Further investigation is needed to determine a single scoring system to be adopted for unified academic reporting and preoperative prediction for the treatment of renal calculi with PCNL.

REFERENCES

1. Scales CD Jr, Smith AC, Hanley JM et al: Prevalence of kidney stones in the United States. Urologic Diseases in America Project. Eur Urol 2012; **62:** 160.
2. Preminger GM, Assimos DG, Lingeman JE et al: Chapter 1: AUA guideline on management of staghorn calculi: diagnosis and treatment recommendations. J Urol 2005; **173:** 1991.
3. Ghani KR, Sammon JD, Bhojani N et al: Trends in percutaneous nephrolithotomy use and outcomes in the United States. J Urol 2013; **190:** 558.
4. Sivalingam S, Cannon ST and Nakada SY: Current practices in percutaneous nephrolithotomy among endourologists. J Endourol 2014; **28:** 524.
5. Mirheydar HS, Palazzi KL, Derweesh IH et al: Percutaneous nephrolithotomy use is increasing in the United States: an analysis of trends and complications. J Endourol 2013; **27:** 979.
6. Thomas K, Smith NC, Hegarty N et al: The Guy's stone score—grading the complexity of percutaneous nephrolithotomy procedures. Urology 2011; **78:** 277.
7. Okhunov Z, Friedlander JI, George AK et al: S.T.O.N.E. nephrolithometry: novel surgical classification system for kidney calculi. Urology 2013; **81:** 1154.
8. Smith A, Averch TD, Shahrour K et al: A nephrolithometric nomogram to predict treatment success of percutaneous nephrolithotomy. J Urol 2013; **190:** 149.
9. Hyams ES, Bruhn A, Lipkin M et al: Heterogeneity in the reporting of disease characteristics and treatment outcomes in studies evaluating

treatments for nephrolithiasis. J Endourol 2010; **24:** 1411.

10. Opondo D, Gravas S, Joyce A et al: Standardization of patient outcomes reporting in percutaneous nephrolithotomy. J Endourol 2014; **28:** 767.
11. Raman JD, Bagrodia A, Bensalah K et al: Residual fragments after percutaneous nephrolithotomy: cost comparison of immediate second look flexible nephroscopy versus expectant management. J Urol 2010; **183:** 188.
12. Raman JD, Bagrodia A, Gupta A et al: Natural history of residual fragments following percutaneous nephrostolithotomy. J Urol 2009; **181:** 1163.
13. de la Rosette JJ, Opondo D, Daels FP et al: Categorisation of complications and validation of the Clavien score for percutaneous nephrolithotomy. Eur Urol 2012; **62:** 246.
14. Andonian S, Okhunov Z, Shapiro EY et al: Diagnostic utility and clinical value of post-percutaneous nephrolithotomy nephrostogram. J Endourol 2010; **24:** 1427.
15. Li R, Louie MK, Lee HJ et al: Prospective randomized trial of three different methods of nephrostomy tract closure after percutaneous nephrolithotripsy. BJU Int 2011; **107:** 1660.
16. Penbegul N, Hatipoglu NK, Bodakci MN et al: Role of ultrasonography in percutaneous renal access in patients with renal anatomic abnormalities. Urology 2013; **81:** 938.
17. Osther PJ, Razvi H, Liatsikos E et al: Percutaneous nephrolithotomy among patients with renal anomalies: patient characteristics and outcomes; a subgroup analysis of the clinical research office of the endourological society global percutaneous nephrolithotomy study. J Endourol 2011; **25:** 1627.
18. Mandal S, Goel A, Kathpalia R et al: Prospective evaluation of complications using the modified Clavien grading system, and of success rates of percutaneous nephrolithotomy using Guy's Stone Score: a single-center experience. Indian J Urol 2012; **28:** 392.
19. Ingimarsson JP, Dagrosa LM, Hyams ES et al: External validation of a preoperative renal stone grading system: reproducibility and inter-rater concordance of the Guy's stone score using preoperative computed tomography and rigorous postoperative stone-free criteria. Urology 2014; **83:** 45.
20. Okhunov Z, Moreira D, George A et al: PD32-09 Multicenter validation of S.T.O.N.E. nephrolithometry. J Urol, suppl., 2014; **191:** e839, abstract PD32-09.
21. Akhavein A, Henriksen C and Bird VG: Prediction of single procedure success rate using S.T.O.N.E. nephrolithometry surgical classification system with strict criteria for surgical outcome. J Urol, suppl., 2013; **189:** e627, abstract 1532.
22. Okhunov Z, Helmy M, Perez-Lansac A et al: Interobserver reliability and reproducibility of S.T.O.N.E. nephrolithometry for renal calculi. J Endourol 2013; **27:** 1303.
23. Kattan MW: Comparison of Cox regression with other methods for determining prediction models and nomograms. J Urol 2003; **170:** S6.
24. Kattan MW: Nomograms are superior to staging and risk grouping systems for identifying high-risk patients: preoperative application in prostate cancer. Curr Opin Urol 2003; **13:** 111.
25. Vicentini FC, Marchini GS, Mazzucchi E et al: Utility of the Guy's stone score based on computed tomographic scan findings for predicting percutaneous nephrolithotomy outcomes. Urology 2014; **83:** 1248.
26. Matlaga BR and Hyams ES: Stones: can the Guy's stone score predict PNL outcomes? Nat Rev Urol 2011; **8:** 363.