Turkish Journal of Emergency Medicine 19 (2019) 91–95

Contents lists available at ScienceDirect

Turkish Journal of Emergency Medicine

journal homepage: www.elsevier.com/locate/tjem

Original Article

A clinical prediction rule for uncomplicated ureteral stone: The STONE score; a prospective observational validation cohort study

Arash Safaie[a,b], Mojdeh Mirzadeh[b], Ehsan Aliniagerdroudbari[c], Sepideh Babaniamansour[d], Alireza Baratloo[a,b,*]

[a] Prehospital Emergency Research Center, Tehran University of Medical Sciences, Tehran, Iran
[b] Department of Emergency Medicine, Sina Hospital, Tehran University of Medical Sciences, Tehran, Iran
[c] School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[d] School of Medicine, Islamic Azad University of Medical Sciences, Tehran, Iran

ARTICLE INFO

Keywords:
Clinical prediction rules
Decision support techniques
Emergency department
Renal colic
Validation studies

ORCIDs:
AS: 0000-0002-1127-6681
MM: 0000-0002-6294-1709
EA: 0000-0003-3797-7977
SB: 0000-0002-4025-5628
AB: 0000-0002-4383-7738

ABSTRACT

Introduction: Renal colic is one of the most common complaints in patients admitted to Emergency Department (ED). Computed Tomography (CT) is the reference standard for the diagnosis of any stones in the kidneys or ureters. However, CT has classical disadvantages, such as radiation exposure, cost and availability. Recently, STONE clinical prediction criteria were suggested to identify uncomplicated ureteral stone cases among patiens admitted to the ED with abdominal pain. Primary objective of this study was the external validation of the STONE criteria.

Methods: This was a diagnostic accuracy study conducted on a prospective, observational cohort. All consecutive patients who underwent a non-enhanced abdominopelvic CT scan in the ED with an initial diagnosis of ureteral stone disease were enrolled. Using a pre-prepared checklist, all data and the final diagnosis according to the CT scan were recorded. STONE score was calculated for all patients. The area under the curve (AUC) of the STONE Score and the CT, the reference standard, were compared using the ROC curve analysis.

Results: Totally, 237 patients (59.9% male) with an average age of 41.54 years (SD: 13.37) were evaluated, and 156 cases (65.8%) were proved to have renal stone. The mean (SD) STONE scores in the groups of patients with renal stone and in the group of patients without renal stone group were 9.1 (2.6) and 6.0 (2.8), respectively ($p < 0.001$). The area under the curve (AUC) for the STONE score was 0.789 (95% confidence interval (CI) 0.725 to 0.852). The optimum threshold value of the STONE score for the diagnosis of a renal stone was 8 or more, which had a sensitivity of 75.0% and a specificity of 70.4%.

Conclusion: Despite the acceptable diagnostic accuracy, further modifications and enhancements of the STONE score are needed to differentiate patients with low risk prior to imaging.

1. Introduction

Renal colic is one of the most common complaints in patients referred to the emergency department (ED). The prevalence of the kidney stones is reported as 1–5% in various areas, which is estimated to be 2–3% in developed countries, and 0.5–1% in developing countries and these stats have an ascendant trend.[1–5] Studies also show that 8–15% of the American and European people will suffer from kidney stones during their lives. Studies in white men showed that at the age of 70, one out of eight people will suffer from this disease.[6] Forty-one percent of the patients with the first renal colic attack will have a second one within 2–3 years, and 75% of patients will have recurrence in 7–10 years.[1,6]

Management of these patients is usually limited to pain control, and in the majority of cases, the stone is expelled by itself without the need for an invasive intervention. The need for an invasive intervention, and the possibility of complications is investigated by the help of radiologic imaging. The reference standard test in suspected ureteral stone is non-enhanced abdominopelvic computed tomography (CT) scan. However, it is time-consuming, expensive, leads to unnecessary radiation exposure, and increases the length of ED stay.[7,8] Numerous efforts have been made to identify uncomplicated ureteral stone cases in the ED in order to reduce unnecessary imaging, reduce costs, and duration of the stay. In this regard, a clinical prediction score named as STONE was

Peer review under responsibility of The Emergency Medicine Association of Turkey.
* Corresponding author. Department of Emergency Medicine, Sina Hospital, Hasanabad Square, Tehran, Iran.
E-mail address: arbaratloo@sina.tums.ac.ir (A. Baratloo).

https://doi.org/10.1016/j.tjem.2019.04.001
Received 20 December 2018; Received in revised form 18 March 2019; Accepted 10 April 2019
Available online 25 April 2019

designed in 2014 by Moore et al., to determine the probability of an uncomplicated ureteral stone presence in patients referred with abdominal pain.[9] During the first phase (derivation cohort) of their study, 1040 patients who had a non-enhanced CT scan with a probability of stone in urinary tract between 2005 and 2010 were retrospectively evaluated with multivariate logistic regression method. Male sex, acute onset of pain, non-black race, presence of nausea or vomiting, and microscopic hematuria were found to be the best factors with highest correlations with stone presence. The value of the STONE clinical prediction criteria and respective risk groups that it predicts were determined as follows: 0–5 low-risk, 6–9 moderate risk, and 10–13 high risk. In the second phase (validation cohort) which was carried out between 2012 and 2013, its validity was confirmed by studying 491 patients.[9]

Due to the novelty of this clinical prediction rule, limited studies have been made to confirm its external validity and the studies reported variety of results provides good background for further studies.[9–14] Therefore, we conducted this survey to study the external validity of the STONE clinical prediction criteria in patients with referred to ED with abdominal pain.

2. Material and methods

2.1. Study design and setting

This prospective observational cohort study was conducted between October 2017 and April 2018 in 3 EDs from educational medical centers affiliated to Tehran University of Medical Sciences (TUMS), Tehran, Iran. The study protocol was approved by ethical committee of TUMS and protocol code of "IR.TUMS.MEDICINE.REC.1396.4804" was assigned to it. In this study, no interventions were performed by the researchers to the patient's diagnostic and treatment processes. The data was collected from the patients who agreed to sign informed consent to this survey and to authorize the use of the data for research purposes. Patient information were anonymized.

2.2. Definitions

Male sex, acute onset of pain, non-black race, presence of nausea or vomiting and microscopic hematuria were defined as the components of the STONE Clinical Prediction Criteria (Table 1).

2.3. Study population

Results of the studies performed by Moore et al. and Daniels et al. were used for the calculation of the sample size.[9,10] The estimated sample size was 233 patients (155 positives and 78 negatives) to find and AUC difference of 0.10 in accuracies as statistically significant with a power of 80% and type 1 error of 5% (PASS version 11, PASS Power Analysis and Sample Size System, NCSS, USA). All consecutive patients who underwent a non-enhanced abdominopelvic CT scan in the ED with an initial diagnosis of ureteral stone disease were enrolled. Patients whose CT image was not visible in the hospital picture archiving and communication system (PACS) for any reason, who did not give a urine sample for dipstick test, and those who left the ED against medical advice before performing a CT scan were excluded from the study. Linguistic, racial and generic limitations were not considered.

Table 1
STONE Clinical Prediction Criteria.

Variable	Points
Sex	
Female	0
Male	2
Duration of pain to presentation	
> 24 hours	0
6–24 hours	1
< 6 hours	3
Race	
Black	0
Non-black	3
Nausea and vomiting	
None	0
Nausea alone	1
Vomiting alone	
Hematuria (on urine dipstick)	
Absent	0
Present	3
Total	0–13

2.4. Data collection

Data was collected by the researchers, since they were merely observers and had no intervention in the process of patient's diagnostic and therapeutic actions. Two emergency medicine specialists with more than 6-years of work experience collected the data. Researchers attended the ED as an observer at different hours of the day, different days of the week, and they filled out a standard data collection form for enrolled patients. Patients' demographic information, starting time of the pain, presence of nausea or vomiting, history of renal stone disease, presence of hematuria in dipstick test, and final diagnosis based on the CT scan were recorded. Two emergency medicine specialists who were blinded to patient's symptoms, evaluated the CTs for the evidence of ureteral stones independently. In case of disagreement, the opinion of a third emergency medicine specialist was used as a tie-breaker. Presence of any stone in an anatomical location between the renal pelvis and ureterovesicular junction with patient reported pain on the same side of the body was interpreted as a complicated renal stone.

2.5. Statistical analysis

The data was manually edited in SPSS software version 23 (IBM corp., USA) and analyzed. Relative risk and odds ratio calculated using Amare software version 2.0 (Safa App, Iran). For each patient the STONE score was calculated. Then, the area under the curve (AUC) of the STONE score and its 95% confidence internal (CI) was calculated using the ROC curve analysis considering non-enhanced abdominopelvic CT scan as the reference standard. Categorical variables were analyzed using chi-square test. Numerical variables were analyzed using independent *t*-test (comparing the means) or Mann-Whitney *U* test (comparing the ranks) based on being parametric or not, respectively. A p-value of 0.05 of less was considered as statistically significant.

3. Results

There were 286 eligible patients, where 237 (%82.9) were enrolled to this study. In the study population, an ureteral stone was present in 156 (65.8%), and was not present in 81 cases (34.2%). The mean (SD) age was 41.54 (13.37) years (min-max: 17–70 years), and 142 of the 237 patients (59.9%) were male.

Comparison of patient's basic data is summarized in Table 2 in order to compute the STONE score in two groups of with stone and without stone. The average age of patients in two groups didn't have significant statistical difference. Also distribution according to gender in two groups was not significantly different but the frequency of male sex in patients with stone was significantly higher than without stone patients ($p < 0.001$). The mean time from the onset of pain until the time of admittance to the ED was significantly in the stone group compared to the group with no stones ($p = 0.001$). The frequency of nausea and vomiting, and presence of RBC in the urine test were both significantly higher in the stone group, as well ($p < 0.05$). The mean STONE score in the stone and no stone groups were 9.1 (SD: 2.6) and 6.0 (SD: 2.8),

Table 2
Comparison of patient's basic information to compute the stone score in patient with stone and without stone.

Variable	Group		p
	With stone (n = 156)	Without stone (n = 81)	
Age (year), mean (SD)	41.4 (14.2)	41.6 (12.9)	0.713
Sex, n (%)			
Male	107 (68.6)	35 (43.2)	< 0.001
Female	49 (31.4)	46 (56.8)	
Pain onset (minute), mean (SD)	33.5 (70.9)	52.4 (65.2)	0.001
Nausea, n (%)			
Yes	113 (72.4)	37 (45.7)	< 0.001
No	43 (27.6)	44 (54.3)	
Vomiting, n (%)			
Yes	76 (48.7)	24 (29.6)	0.005
No	80 (51.3)	57 (70.4)	
RBC presence in urine dipstick, n (%)			
Yes	92 (59.0)	10 (12.3)	< 0.001
No	64 (41.0)	71 (87.7)	
STONE score, mean (SD)	9.1 (2.6)	6.0 (2.8)	< 0.001

respectively, which was found to be statistically significant ($p < 0.001$).

Fig. 1 shows the ROC curve of the STONE score according to presence of stone for studied patients. The area under the curve (AUC) of the STONE score for the diagnosis of a renal stone was 0.789 (95% confidence interval 0.725 to 0.852). The optimal threshold value of the STONE score for the diagnosis of a stone was greater than or equal to 8, with a sensitivity of 75% and a specificity of 70.4%. Positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR+), negative likelihood ratio (LR-) and accuracy were 82.9%, 60.2%, 2.53, 0.355 and 74.26, respectively.

The frequencies of the stone presence in low, moderate and high risk groups accordin to STONE score in this study was 31.7% (n = 19/60), 65.6% (n = 61/93) and 90.5% (n = 76/84), respectively. Fig. 2 shows the number of patients in each STONE risk group stratified according to presence of stones. There was a statistically significant difference

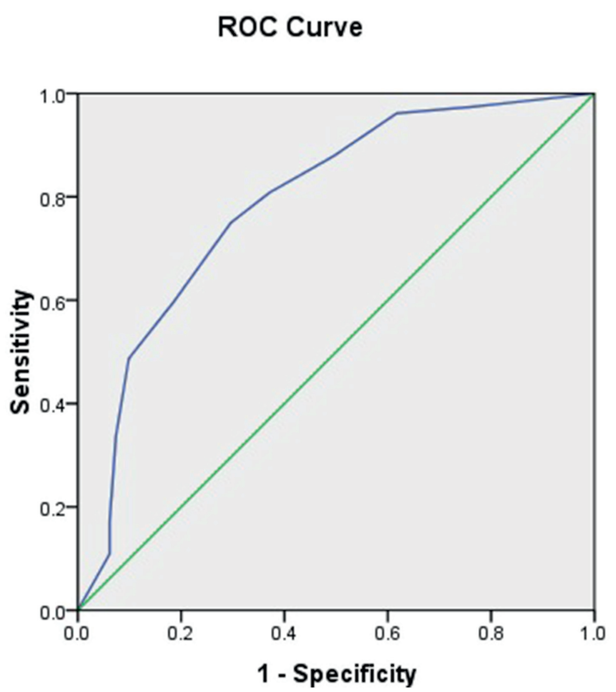

Fig. 1. ROC curve of the STONE score for a renal stone in studied patients.

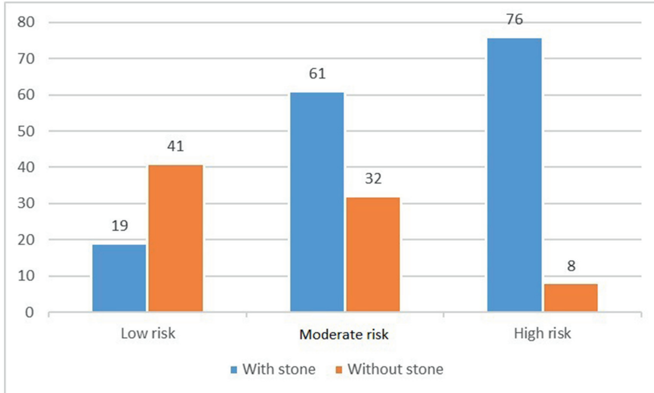

Fig. 2. Number of patients with and without stone in each STONE risk group.

between the frequencies of stone presence among STONE risk groups (chi-squared, $p < 0.0001$). Proportion of patients with a renal stone was significantly higher in the STONE high risk group compared to low and moderate risk groups ($p < 0.001$). Relative risk of a renal stone was 2.86 (95% CI: 1.96 to 4.17) in the STONE high risk, and 2.07 (95% CI: 1.39 to 3.10) in the STONE moderate risk groups compared to low risk group, and both were statistically significant ($p < 0.005$).

Diagnoses in the no stone group were gastroenteritis, Mittelschmerz, menstrual pain, uncomplicated urinary tract infection, and biliary colic. No patients were diagnosed with a critical diagnosis or required emergent surgical intervention. Details of those diagnoses in the group of patients with a negative CT scan were listed in Table 3.

4. Discussion

Current study aimed to investigate the external validity of the STONE score in patients referred to the ED with probability of ureteral stone. According to the results, the mean and the standard deviation of the STONE score in the group with stone (9.1) were significantly higher than the patients without a stone (6.0).

Either in the original study by Moore et al. in which the STONE clinical prediction rule was designed, or in the present study, only patients evaluated with non-enhanced abdominopelvic CT, as it was accepted as the reference standard for the diagnosis of ureteral stones, were included. This method is sensible and already preferred in many other similar studies. The three primary predictors of stone passage without the need for surgical intervention are calculus size, location, and degree of patient's pain. The most important factor that relates to passage of a calculus though the genitourinary tract is its size. Apparently there is no other diagnostic method, except CT for size measurement; so, it has become a routine approach for all suspected patients in most of academic centers such as our study. [11]To the best of our knowledge, there are no criteria defined to differentiate low versus high risk patients, and usually decision was reached according to the patients' history and physical exam, and if the possibility of kidney stone is raised, a CT scan is requested for the patient.

Table 3
Diagnoses in the group of patients with a negative CT scan.

Diagnosis	Male (n = 35)	Female (n = 46)
	Number (%)	
Gastroenteritis	4 (11.4)	1 (2.2)
Mittelschmerz	–	6 (13.0)
Menstrual pain	–	9 (19.6)
Uncomplicated urinary tract infection	9 (25.7)	7 (15.2)
Biliary colic	13 (37.1)	18 (39.1)
Unknown	9 (25.7)	5 (10.9)

Table 4
The results of the validation studies on STONE criteria.

Study	Sample size	Sampling	Low	Moderate	High	AUC
Moore et al., 2014	1040	retrospective	8.3%	51.6%	89.6%	0.82
Moore et al., 2014	491	prospective	9.2%	51.3%	88.6%	0.79
Schoenfeld et al., 2015	134	retrospective	13.3%	58.0%	89.7%	0.87
Daniels et al., 2016	835	prospective	10.77%	44.53%	87.50%	N/R
Hernandez et al., 2016	536	retrospective	14.0%	48.3%	75.8%	N/R
Kim et al., 2016	700	retrospective	21.8%	80.1%	98.7%	0.92
Wang et al., 2016	845	retrospective	13.5%	32.2%	72.7%	0.78

AUC: area under the curve; N/R: not reported.

In the literature, there are five other studies present aimed to determine the external validity of the STONE score, besides the original study by Moore at al. The results are summarized in Table 4.

In the validation study conducted by Moore et al. 491 patients were evaluated. The AUC was 0.792 (95% CI: 0.756 to 0.828). In 1.6% of high risk patients, there was another acutely important alternative finding beside ureteral stone.[9] Findings of our study are consistent with the study of Moore et al., however, in our study the AUC and diagnostic accuracy was slightly lower. The reason for this difference may be due to the difference in the sample size of two studies and the difference in the demographic indices of individuals in the studies. Comparison of the findings of current study and Moore's is shown in Fig. 3.

Daniels and colleagues studied 835 patients in a prospective observational study. In addition to evaluating the STONE score, patients also underwent an ultrasonic examination to evaluate the presence of hydronephrosis. The impact of adding presence of hydronephrosis was evaluated and it resulted in a slight increase in sensitivity among low and moderate risk groups. It also improved specificity from 67% to 98%, and 42% to 98% in low and moderate risk groups, respectively. However, the presence of hydronephrosis didn't change the sensitivity and specificity of the STONE score in high risk patients. They also used the findings of 491 studied patients to validate the STONE criteria and reported similar results with Moore et al. in all three risk groups.[10]

Hernandez et al. in a retrospective study investigated 536 patients in order to confirm the external validity of the STONE score.[12] The results of our study regarding the diagnostic accuracy of the STONE score in high risk patients is consistent with the findings of Hernandez's, but differ in respect of overall diagnostic accuracy.

While examining external validity of the STONE score, Kim and colleagues also have been trying to create a new rule in this field. They retrospectively studied 700 patients that 555 (79%) of them had a ureter stone. The AUC of the STONE score was 0.92, and the sensitivity for a high risk STONE score was 0.56. By removing nausea, vomiting and racial properties, and replacing them with the presence of a CRP level less than 0.5, the AUC increased to 0.94 and sensitivity to 0.80.[13]

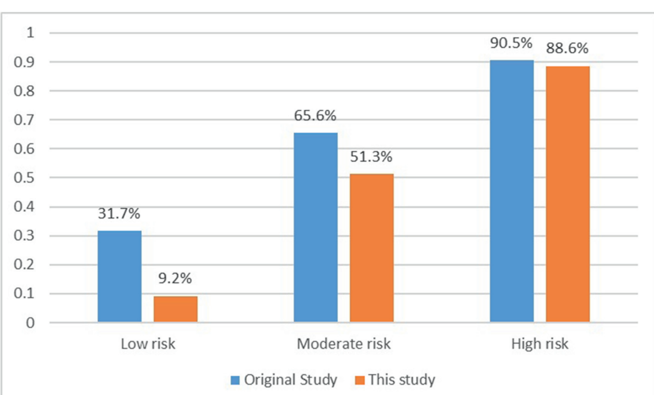

Fig. 3. Prevalence of renal stone in the original validation study, and our study stratified according to STONE score risk groups.

The results of the Kim's study are similar to the findings of our study. However, in our study the AUC, therefore the diagnostic accuracy, was modestly lower. The reason for this variation may be due to the difference in the sample size, the difference in the demographic indices of individuals and the difference in the implementation method.

Schoenfeld et al., in a retrospective study evaluated the external validity of the STONE score in patients aged 18 to 50 years. Of the 134 patients who were included into the study, 56.7% were women, the average age was 37 years, and 52% had a kidney stone causing an obstruction. The AUC of the STONE score was 0.87, and the sensitivity and specificity of a STONE score of 8 or more for the diagnosis of renal stone was 78.6% and 84.4%, respectively.[14] The results of this study are consistent with the findings of our study.

In a retrospective study, Wang et al. evaluated the external validity of the STONE score, and 331 of the 845 patients participated in the study (39%) had ureteral stone. The overall accuracy of the STONE score was higher than the physician's gestalt (the AUC of 0.78 versus 0.68). The presence of stone in low, moderate and high risk groups were 14%, 32% and 73%, respectively. The sensitivity and specificity of a high score for the presence of a stone were respectively 53% and 87%.[15] The results of the study are equivalent with the findings of our study. However, in our study the sensitivity achieved was 75%, which was higher than the aforementioned study.

As discussed above, all previous studies, and the current study, showed that the STONE score successfully stratified patients into three consecutive risk groups (low, moderate and high) for ureteral stone, and it was more accurate than the physician's prediction. There was a slight difference between the results of the studies, which may be due to the difference in sample sizes and different demographic indices. For instance, although racial status is one of the items in the STONE scoring system, there were no patients from the black race in our study, which was due to the topological state of the centers where the study was performed. Moreover, in comparison with the original study, the sample size of our study was lower. These observations suggest that in order to defer a CT scan for a patient referred to the emergency department with a complaint of renal colic, the STONE score can help physicians to have a more effective decision making with lesser harms for the patients. Nevertheless, further development of the STONE score is needed to produce a more successful decision rule.

4.1. Limitations

First, although we prospectively evaluated a smaller sample size compared to previous studies, it seems that the kappa coefficient of the STONE score is moderate, and may be improved by adding other predictors estimating the presence of a stone in future studies to achieve a better scoring system for the estimation of stone disease. Second, despite being rare, those who did not have a CT scan (decided by the physician) for evaluation were not included in this study, which may have led to a work-up bias. Third, short-term follow-up was also limited to the hospital admission period, and this is the most important limitation of the current study. Fourth, we have no information on the readmission, superinfection, stone passage, or further interventions. Last, STONE score has a race-dependent component, and non-black race gets 3 points. However, there were no black patients present in this study. This is another important limitation, since all non-black patients had an initial score of 3 points at baseline.

5. Conclusion

Based on the results obtained from this study, although the diagnostic accuracy of the STONE score is acceptable, the STONE score needs to be modified in order for it to differentiate patients with low risk prior to imaging.

Author contribution statement

1) conceived and designed the experiments: AS, AB
2) performed the experiments: MM, EA, SB
3) analyzed and interpreted the data: AS, MM
4) contributed reagents, materials, analysis tools or data: AS, MM, EA, SB, AB
5) wrote the paper: EA, SB, AB

All the authors read the final version and approved the details.

Conflict of interest statement

The authors declare no conflict of interest.

Funding

No financial support has been received for this project.

Disclaimer

None.

Acknowledgments

The authors wish to express their appreciation and thanks to all colleagues for their cooperation in the present study. It should be noted that the present article was extracted from the thesis of Dr. Mojdeh Mirzadeh to achieve her specialist degree in Emergency Medicine from Tehran University of Medical Sciences.

References

1. Lee YH, Huang WC, Tsai JY, et al. Epidemiological studies on the prevalence of upper urinary calculi in Taiwan. *Urol Int.* 2002;68:172–177.
2. Shoag J, Tasian GE, Goldfarb DS, et al. The new epidemiology of nephrolithiasis. *Adv Chron Kidney Dis.* 2015;22:273–278.
3. Edvardsson VO, Indridason OS, Haraldsson G, et al. Temporal trends in the incidence of kidney stone disease. *Kidney Int.* 2013;83:146–152.
4. Yasui T, Iguchi M, Suzuki S, et al. Prevalence and epidemiological characteristics of urolithiasis in Japan: national trends between 1965 and 2005. *Urology.* 2008;71:209–213.
5. Fukuhara H, Ichiyanagi O, Kakizaki H, Naito S, Tsuchiya N. Clinical relevance of seasonal changes in the prevalence of ureterolithiasis in the diagnosis of renal colic. *Urolithiasis.* 2016;44:529–537.
6. Hess B. Pathophysiology, diagnosis and conservative therapy in calcium kidney calculi. *Ther Umsch.* 2003;60:79–87.
7. Westphalen AC, Hsia RY, Maselli JH, et al. Radiological imaging of patients with suspected urinary tract stones: national trends, diagnoses, and predictors. *Acad Emerg Med.* 2011;18:699–707.
8. Hyams ES, Korley FK, Pham JC, et al. Trends in imaging use during the emergency department evaluation of flank pain. *J Urol.* 2011;186:2270–2274.
9. Moore CL, Bomann S, Daniels B, et al. Derivation and validation of a clinical prediction rule for uncomplicated ureteral stone–the STONE score: retrospective and prospective observational cohort studies. *BMJ.* 2014;348:g2191.
10. Daniels B, Gross CP, Molinaro A, et al. STONE PLUS: evaluation of emergency department patients with suspected renal colic, using a clinical prediction tool combined with point-of-care limited ultrasonography. *Ann Emerg Med.* 2016;67:439–448.
11. Schoenfeld EM, Pekow PS, Shieh MS, et al. The diagnosis and management of patients with renal colic across a sample of US hospitals: high CT utilization despite low rates of admission and inpatient urologic intervention. *PLoS One.* 2017;12:e0169160.
12. Hernandez N, Song Y, Noble VE, et al. Predicting ureteral stones in emergency department patients with flank pain: an external validation of the STONE score. *World J Urol.* 2016;34:1443–1446.
13. Kim B, Kim K, Kim J, et al. External validation of the STONE score and derivation of the modified STONE score. *Am J Emerg Med.* 2016;34:1567–1572.
14. Schoenfeld EM, Poronsky KE, Elia TR, et al. Validity of STONE scores in younger patients presenting with suspected uncomplicated renal colic. *Am J Emerg Med.* 2016;34:230–234.
15. Wang RC, Rodriguez RM, Moghadassi M, et al. External validation of the STONE score, a clinical prediction rule for ureteral stone: an observational multi-institutional study. *Ann Emerg Med.* 2016;67:423–432.