External validation of STONE, modified STONE, and CHOKAI scores for the diagnosis of ureteral stones in the Turkish population

Yahya Ayhan Acar1 and Emin Uysal2

Abstract
Background: Scoring systems seem to be effective in the management of patients with uncomplicated ureteral stones. However, their efficiency may differ by population.
Objectives: We aimed to validate STONE, modified STONE, and CHOKAI scores for the diagnosis of ureteral stones in the Turkish population.
Methods: We conducted a retrospective chart review between 01 February 2018 and 30 November 2018, in an academic emergency department. Demographics, laboratory findings, and radiologic tests of patients with flank pain were obtained. Computed tomography was used as the gold standard for the diagnosis of ureteral stones. STONE, modified STONE, and CHOKAI scores were calculated for each patient. The performance of the scoring systems was compared in terms of their specificity, sensitivity, positive likelihood ratio, negative likelihood ratio, negative predictive value, and positive predictive value.
Results: A total of 157 patients were included in the study. The mean age was 38.47 ± 14.87 years, and 103 (65.6%) of the patients were males. The prevalence of ureteral stones was 84.0%, 88.9%, and 85.0% in the high-risk patients and 12%, 9.4%, and 22.7% in the low-risk patients for the STONE, modified STONE, and CHOKAI scores, respectively. Area under the curve values for the STONE, modified STONE, and CHOKAI scores were 0.776 (p = 0.001; 0.692–0.860 95% confidence interval), 0.825 (p < 0.001; 0.749–0.901 95% confidence interval), and 0.869 (p < 0.001; 0.806–0.932 95% confidence interval), respectively. The specificity and sensitivity values of STONE, modified STONE, and CHOKAI scores for the diagnosis of ureteral stones were 64.71, 71.70; 70.59, 87.74; and 66.67, 90.57, respectively.
Conclusion: The CHOKAI score displayed the best performance compared to STONE and modified STONE in diagnosing ureteral stones in the Turkish population.

Keywords
Diagnosis, emergency medicine, ureter, urinary calculi

Introduction
Flank pain is one of the most prevalent causes of emergency department (ED) admission. The lifetime expectancy of urolithiasis in the population is 5%, of which 8% of emergency visits require admission.1-3 All patients must be assessed carefully according to their history, physical examination, and laboratory findings. Urinary stones are frequently treated in the ED, and patients are usually evaluated with computed tomography (CT).

CT is the gold standard for the diagnosis of urolithiasis. Patients with known kidney disease, history of malignancy, infection findings (fever or the presence of leukocytes on urine analysis), or a previous urological procedure (including

1Department of Emergency Medicine, Gulhane School of Medicine, University of Health Sciences, Ankara, Turkey
2Department of Emergency Medicine, Bagcilar Training and Research Hospital, Istanbul, Turkey

Corresponding author:
Yahya Ayhan Acar, Department of Emergency Medicine, Gulhane School of Medicine, University of Health Sciences, Etlik, 06010 Ankara, Turkey.
Email: yahyaacar@gmail.com
lithotripsy or ureteral stents) are likely to undergo CT. However, it has not been shown to improve patient outcomes for uncomplicated cases because most kidney stones are benign and will pass spontaneously.\(^4,5\) Kidney stones have a high recurrence rate, especially in younger patients. Therefore, performing CT for every emergency and urology clinic admission may increase the risk of malignancy in the long term. Several risk stratification and scoring systems for the diagnosis of urolithiasis have been developed to help clinicians in the management of these patients. These scoring systems have been implied not only for the diagnosis but also to reducing the radiation dose burden from CT and the cost of treatment per patient.\(^6\) Moreover, additional imaging will result in an increased length of stay time and increased costs. These objective clinical scoring systems for ureteral stones may assist emergency physicians in decision-making and allow them to manage uncomplicated patients without imaging.

None of the current scoring systems for ureteral stones has been shown to be the gold standard in primary and validation studies. One of the most studied scoring systems is the STONE protocol, which was proposed by Moore et al.\(^7\) This protocol uses information on sex, duration of the pain, race, presence of nausea and vomiting, and hematuria on urinalysis.\(^7\) Recently, it was advocated that diagnosis based on the STONE scoring system might reduce the need for CT in the diagnosis of ureteral stones.\(^8-10\) However, the universal application of the STONE scoring system seems restricted because “race” cannot be quantified in relatively homogeneous populations. Therefore, enhancements were suggested.\(^11\) Kim et al.\(^12\) proposed a modified STONE score for the Korean population, while Fukuhara et al.\(^13\) proposed the CHOKAI score for the Japanese population. In addition, a suggestion to include point of care ultrasonography (US) in the STONE scoring system was made. The STONE PLUS scoring system was developed by Daniels et al.\(^14\)

To the best of our knowledge, no study has investigated the validity of these scoring systems in the Turkish population. The present study aimed to compare the accuracy of the three different scoring systems for the diagnosis of ureteral stones in the Turkish population.

**Methods**

This is a retrospective descriptive study. The study protocol was approved by the ethical committee of the Bagcilar Training and Research Hospital (approval number: 2019.04.2.02.017.r3.040). Written informed consent was not necessary because no patient data have been included in the article. The study was conducted at the Department of Emergency Medicine of Bagcilar Training and Research Hospital, Istanbul, Turkey, between 01 February 2018 and 30 November 2018. The host institution is a tertiary care center with 1300 daily emergency admissions.

**Study population and data collection**

All the patients with flank pain admitted to the ED were screened for eligibility using the data in the hospital information system and patient charts. Demographics (age and gender) and history (known history of urinary stone, duration of the pain, nausea, and vomiting) were recorded. Laboratory findings such as a urinalysis, kidney function tests, and infection markers (including white blood cell (WBC) count, neutrophil count, and C-reactive protein (CRP) levels) were recorded. Urinary tract infection was defined as the presence of leukocytes on urinalysis. In addition, if performed, reports from radiological examinations, including plain radiographs, US, and CT were evaluated. In the hosting hospital, all the US (Toshiba Aplio™ 300, Canon Medical Systems, Tokyo, Japan) and CT (Ingenuity Core128, Philips Inc., Netherlands) examinations are reported by a radiologist, and these reports were used for the study. CT reports were considered the gold standard for diagnosing urolithiasis. Patients with findings that were incompatible with ureteral stone underwent further evaluation to investigate differential diagnoses according to the routine clinical policy of the hosting institution.

Patients were excluded if they were under 18 years old, pregnant, had flank pain associated with trauma, urinary tract infection, were unable to speak, suffered a loss of consciousness, had malignancy, or unstable vital signs. STONE, modified STONE, and CHOKAI scores were calculated for each patient according to the original reports, and the details are presented in Table 1.

**Statistical analysis**

Descriptive statistics are presented as frequency, percentage(%), and mean ± standard deviation (SD). The distribution of the data was assessed with the Kolmogorov–Smirnov test. A chi-square test was used to compare the categorical variables. The performance of the different scoring systems was interpreted using the area under the curve (AUC) of the receiver operating characteristics (ROC) and by calculating the specificity, sensitivity, positive likelihood ratio (LR+), negative likelihood ratio (LR−), positive predictive value (PPV), and negative predictive value (NPV). The results were separated into those for high- and low-risk groups, according to the cut-off values derived from the ROC analysis for the three scoring systems. All statistical tests were performed with the Predictive Analytics Software (PASW, version 18; SPSS Inc., Chicago, IL, United States).

**Results**

A total of 409 patients were reviewed. However, only 157 met the inclusion criteria and were included in the study. A flowchart of the study is shown in Figure 1, and the demographics
of the patients are presented in Table 2. There was no difference between the urolithiasis and non-urolithiasis groups in terms of age and CRP levels (p=0.585 and 0.077, respectively). However, there was a significant difference in terms of gender, kidney stone history, nausea/vomiting history, duration of pain, hematuria, hydronephrosis on US (p=0.012, <0.001, 0.022, 0.006, <0.001, <0.001, respectively). No urolithiasis patient required hospital admission or emergent
urologic intervention. There were no patients of a non-Caucasian origin. Therefore, all the patients were assigned three points in the STONE scoring system in the race category.

According to the STONE scoring system, in the high-risk group, 84.0% of the patients had ureteral stones, while 12.0% of the patients in the low-risk group had ureteral stones (Figure 2).

ROC curves are presented in Figure 3. The AUC values for the STONE, modified STONE, and CHOKAI scores were 0.776 (p = 0.001; 0.692–0.860 95% confidence interval (CI)), 0.825 (p < 0.001; 0.749–0.901 95% CI), and 0.869 (p < 0.001; 0.806–0.932 95% CI), respectively.

During the performance analysis, the CHOKAI scoring system performed better than the STONE and modified STONE scoring systems. The specificity values of the STONE, modified STONE, and CHOKAI scores were 64.71, 70.59, and 66.67, respectively, whereas sensitivity values were 71.70, 87.74, and 90.57, respectively (Table 3).

In the non-ureteral stone group (n = 51), 39 patients did not have a definitive diagnosis. Among the 12 patients with a definitive diagnosis, 4 were diagnosed with mesenteric lymphadenitis, 3 with acute appendicitis, 1 with abdominal aortic aneurysm, 1 with lower lobe pneumonia, 1 with inflammatory bowel disease, 1 with a dermoid cyst, and 1 with adrenal adenoma. Risk stratifications according to the three scoring systems are presented in Table 4.

### Table 2. Demographics of the subjects according to parameters of STONE, modified STONE, and CHOKAI scores.

| Parameter                  | Total n (%) | Urolithiasis positive n (%) | Urolithiasis negative n (%) | p* |
|----------------------------|-------------|-----------------------------|-----------------------------|----|
| Gender Male                | 103 (65.6)  | 77 (49.0)                   | 26 (16.6)                   | 0.012 |
| Female                     | 54 (34.4)   | 29 (18.5)                   | 25 (15.9)                   | 0.585 |
| Age ≥60 years              | 16 (10.2)   | 12 (8.6)                    | 4 (2.6)                     | 0.585 |
| <60 years                  | 141 (89.8)  | 94 (59.9)                   | 47 (29.9)                   |     |
| Origin Caucasian          | 157 (100.0) | 106 (67.5)                  | 51 (32.5)                   | N/A |
| Non-Caucasian None         | None        | None                        | None                        |     |
| History of kidney stone    |             |                             |                             |     |
| Positive                   | 89 (56.7)   | 82 (52.2)                   | 7 (4.5)                     | <0.001 |
| Negative                   | 68 (43.3)   | 24 (15.3)                   | 44 (28.0)                   |     |
| Nausea and vomiting history|             |                             |                             |     |
| Nausea only                | 65 (41.4)   | 40 (25.5)                   | 25 (15.9)                   | 0.022 |
| Nausea and vomiting        | 44 (28.0)   | 37 (23.6)                   | 7 (4.4)                     |     |
| None                       | 48 (30.6)   | 29 (18.5)                   | 19 (12.1)                   |     |
| Duration of the pain       |             |                             |                             |     |
| <6 h                       | 65 (41.4)   | 51 (32.5)                   | 14 (8.9)                    | 0.006 |
| 6 h to 1 day               | 41 (26.1)   | 29 (18.5)                   | 12 (7.6)                    |     |
| >1 day                     | 51 (32.5)   | 26 (16.6)                   | 25 (15.9)                   |     |
| Hematuria                  |             |                             |                             |     |
| Positive                   | 108 (68.8)  | 89 (56.7)                   | 19 (12.1)                   | <0.001 |
| Negative                   | 49 (31.2)   | 17 (10.7)                   | 32 (25.5)                   |     |
| Hydronephrosis on ultrasound|             |                             |                             |     |
| Positive                   | 62 (39.5)   | 58 (36.9)                   | 4 (2.6)                     | <0.001 |
| Negative                   | 95 (60.5)   | 48 (30.6)                   | 47 (29.9)                   |     |
| C-reactive protein <0.5 mg/dL | 29 (18.5) | 24 (15.3)                   | 5 (3.2)                     | 0.077 |
| ≥0.5 mg/dL                 | 128 (81.5)  | 82 (52.2)                   | 46 (29.3)                   |     |

N/A: non-applicable.

*Chi-square test.

Figure 2. Prevalence of ureteral stones according to the STONE, modified STONE, and CHOKAI scores.

Discussion

The present study showed that the STONE, modified STONE, and CHOKAI scoring systems are valid for diagnosing ureteral stones in the Turkish population. In daily practice, history and physical examination findings are essential, but emergency physicians can use scoring systems as a complementary tool. Tests with high sensitivity are preferred as the exclusion test in cases with a low pre-test probability. Therefore, among STONE, modified STONE, and CHOKAI scores, the one with a higher sensitivity should be preferred. CHOKAI score showed the highest sensitivity among the scoring systems evaluated in our study, and
emergency physicians may prefer it. However, no test had 100% sensitivity, and it may be beneficial to modify these tests or develop a different scoring system.

In a retrospective study, Turk and Un15 reported that the male sex, presence of hematuria, family history of ureteral stones, nausea, and emesis were predictive factors for urolithiasis in the Turkish population. These parameters are included in all three scoring systems, and our findings are relatively consistent with those of previous studies.

Hernandez et al.16 conducted a study for the external validation of the STONE score. In their study, the low-risk group had higher numbers (24.1%) compared to the original study (8.3%–9.2%), and they concluded that the high prevalence of ureteral stones in the low-risk group should be investigated. In the current study, the prevalence of urinary stones was over 9.4% in the low-risk group and over 70% in the moderate-risk group. The authors concluded that these scoring systems should be evaluated, especially in the low- and moderate-risk groups. In addition, experienced emergency physicians may predict urolithiasis without a radiological examination. However, the STONE scoring system was reported to be more precise than physician gestalt.9 The current study did not assess physician gestalt. However, the lower prevalence of ureteral stones in the low-risk group might be attributed to high physician gestalt.

Cochon et al.17 reported that CT in high-risk patients was not advantageous compared to the STONE scoring system. In addition, assessing hydronephrosis in low-, and moderate-risk patients resulted in a modest improvement of the STONE scoring system, but in high-risk patients, renal US did not alter the performance of the STONE scoring system.14 Our results showed that scoring systems are valid, especially in the high-risk stratified population, and these findings are consistent with literature. Adding US to scoring systems seems as an essential discussion point and needed to be detailed. US was first applied in the STONE PLUS, but it did not cause a marked increase in the performance of the STONE system.14 Then, US was included in

**Table 3.** Sensitivity, specificity, PPV, NPV, LR+, and LR− at the optimal cut-off value of 8 for the STONE score, 7 for the modified STONE score, and 6 for the CHOKAI score.

|          | Sensitivity (CI) | Specificity (CI) | LR+ (CI) | PPV (CI) | NPV (CI) |
|----------|------------------|------------------|----------|----------|----------|
| STONE    | 71.70 (62.12–80.02) | 64.71 (50.07–77.57) | 2.03 (1.37–3.00) | 80.85 (74.08–86.19) | 52.38 (43.31–61.30) |
| Modified STONE | 87.74 (79.94–93.31) | 70.59 (56.17–82.51) | 2.98 (1.94–4.59) | 86.11 (80.11–90.51) | 73.47 (61.76–82.60) |
| CHOKAI   | 90.57 (83.33–95.38) | 66.67 (52.08–79.24) | 2.72 (1.83–4.02) | 84.96 (79.22–89.32) | 77.27 (64.63–86.35) |

CI: 95% confidence interval; LR+: positive likelihood ratio; LR−: negative likelihood ratio; PPV: positive predictive value; NPV: negative predictive value.
the CHOKAI scoring system, which emerged as a more sensitive system than STONE. The main difference between these two scoring systems is that STONE PLUS uses point-of-care US, while CHOKAI uses routine US. In our study, routine US findings were used to compare our results with CHOKAI. We suggest including US in the scoring systems. However, the assessment method should be explained clearly (e.g. by radiologist or emergency physician; for only hydronephrosis or another measurement).

The CHOKAI scoring system is a novel risk stratification system for ureteral stones. The CHOKAI scoring system was reported to perform better in the diagnosis of ureteral stones than the STONE scoring system. The CHOKAI scoring system has no “race” criteria. However, “medical history of ureteral stones” is included. The current study evaluated the CHOKAI scoring system, and it displayed the best performance in the Turkish population. These findings suggest the need to reassess the “race” item in scoring systems, especially when evaluating relatively homogeneous populations. To the best of our knowledge, no other study has compared the CHOKAI and STONE scoring systems in homogeneous populations.

In a validation study of the STONE scoring system, Schoenfeld et al. reported that the STONE scoring system is valid in young subjects, and the mean age of their study population was 37 years. In the current study, the mean age is 38 years, and the STONE protocol worked effectively, particularly in the high-risk group. The reduced performance in the low- and moderate-risk groups might be attributed to the “race” item, which was negative for all the subjects in our study.

Limitations

The main limitation of the current study is the retrospective design and that US was not performed on all the patients. The generalizability of the findings is restricted because of the single-center design of the study. In addition, we did not assess physician gestalt.

The CHOKAI scoring system displayed the best performance for diagnosing ureteral stones in the Turkish population among the three scoring systems reviewed. The STONE scoring system may not work universally, especially in populations with few or no citizens of non-Caucasian origin. The authors conclude that the STONE, modified STONE, and CHOKAI scoring systems are valid in the Turkish population. However, none of the scoring systems reviewed performed flawlessly.

Author contributions

Y.A.A. contributed to the design of the work, the interpretation of data, and drafting; E.U. contributed to the acquisition of data, revising the manuscript critically for important intellectual content, and final approval of the version to be published.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Availability of data and materials

Data were submitted through the submission system.

Informed consent

Written informed consent was waived by ethical committee, because no patient data have been included in the article.

Ethical approval

The study protocol was approved by the ethical committee of Bagcilar Training and Research Hospital, Istanbul, Turkey (approval number: 2019.04.2.02.017.r3.040).

Table 4. Definitive diagnoses in the ureter stone negative group and risk stratification in three scoring systems.

| Definitive diagnosis                  | Risk stratification in STONE scoring system | Risk stratification in modified STONE scoring system | Risk stratification in CHOKAI scoring system |
|---------------------------------------|--------------------------------------------|-----------------------------------------------|---------------------------------------------|
| Mesenteric lymphadenitis-1            | Moderate                                   | Low                                           | Low                                         |
| Mesenteric lymphadenitis-2            | Moderate                                   | Low                                           | Low                                         |
| Mesenteric lymphadenitis-3            | Moderate                                   | Low                                           | Low                                         |
| Mesenteric lymphadenitis-4            | Low                                        | Low                                           | Low                                         |
| Acute appendicitis-1                  | Low                                        | Low                                           | Low                                         |
| Acute appendicitis-2                  | Low                                        | Low                                           | Low                                         |
| Acute appendicitis-3 (with free fluid in pelvic region) | High                                      | Moderate                                      | Low                                         |
| Abdominal aortic aneurysm             | High                                       | High                                          | High                                        |
| Lower lobe pneumonia                  | Low                                        | Low                                           | Low                                         |
| Inflammatory bowel disease            | Moderate                                   | Low                                           | Low                                         |
| Dermoid cyst                          | Moderate                                   | Low                                           | Low                                         |
| Adrenal adenoma                       | Low                                        | Low                                           | Low                                         |
Human rights statement
Authors declare that the work has been conducted in full accordance with the ethical standards on human subjects as well as with the Helsinki Declaration.

ORCID iD
Yahya Ayhan Acar https://orcid.org/0000-0002-9572-5738

References
1. Elder JW, Delgado MK, Chung BI, et al. Variation in the intensity of care for patients with uncomplicated renal colic presenting to U.S. emergency departments. J Emerg Med 2016; 51(6): 628–635.
2. Eaton SH, Cashy J, Pearl JA, et al. Admission rates and costs associated with emergency presentation of urolithiasis: analysis of the nationwide emergency department sample 2006-2009. J Endourol 2013; 27(12): 1535–1538.
3. Yang AW, Johnson JD, Fronczak CM, et al. Lunar phases and emergency department visits for renal colic due to ureteral calculus. PLoS ONE 2016; 11(6): e0157589.
4. Villa L, Giusti G, Knoll T, et al. Imaging for urinary stones: update in 2015. Eur Urol Focus 2016; 2(2): 122–129.
5. Innes GD, Scheuermeyer FX, Law MR, et al. Sex-related differences in emergency department renal colic management: females have fewer computed tomography scans but similar outcomes. Acad Emerg Med 2016; 23(10): 1153–1160.
6. 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(1): e0169160.