GMR
The Original

COMPARISON OF DIAGNOSTIC ACCURACY OF CHOKAI AND STONE SCORES TO DIAGNOSE URETERIC CALCULI

Shafeullah Ahadi[1]*, Atif Javed[2], Muhammad Farjad[3], Quratulain Badar[4], Muhammad Waqas Arshad[5], Muhammad Owais Ramzan[6], Azhar Ghayoor Yousaf[7]

[1]PGR MS, Department of Urology, Dr Zaiuddin University Hospital, Karachi, Pakistan
[2]Associate professor, Department of Urology, Dr Zaiuddin University Hospital, Karachi, Pakistan
[3]Consultant urology, Department of Urology, Dr Zaiuddin University Hospital, Karachi, Pakistan
[4]Assistant professor /consultant radiation oncologist, Department of Oncology, Dr Zaiuddin University Hospital, Karachi, Pakistan
[5]Registrar Urology, Department of Urology, Dr.Ziuddin university Hospital, Karachi, Pakistan
[6]M.S Postgraduate Resident, Department of Urology, Dr Ziauddin University Hospital, Karachi, Pakistan
[7]PGR, Department of Urology, Dr Zaiuddin University Hospital, Karachi, Pakistan
***Correspondence Author:** Dr Shafeullah Ahadi, Email: ahadishafe@gmail.com

ABSTRACT

Objective: To compare the diagnostic accuracy of the CHOKAI and STONE scores for detecting ureteric calculi, using non-contrast CT-KUB as the reference standard, in a Pakistani emergency department setting.

Methods: This cross-sectional study was conducted at the Department of Urology, Ziauddin Hospital, Karachi. A total of 231 adult patients (aged ≥18 years) presenting with acute renal colic were consecutively enrolled. CHOKAI and STONE scores were calculated for all participants, followed by non-contrast CT-KUB imaging. Diagnostic performance parameters including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the receiver operating characteristic curve (AUC) were calculated. Subgroup analyses were performed by stone size (<3 mm vs ≥3 mm) and BMI category (≤30 kg/m² vs >30 kg/m²).

Results: Among 231 patients (median age= 38.0 years), non-contrast CT-KUB confirmed ureteric calculi in 170 (73.6%). At the pre-specified threshold of ≥10, the CHOKAI score showed higher sensitivity than the STONE score (82.9% vs 17.1%) but lower specificity (42.6% vs 90.2%). Overall diagnostic accuracy was superior for CHOKAI (72.3% vs 36.4%, p<0.001). The CHOKAI score achieved a significantly higher area under the receiver operating characteristic curve than the STONE score (AUC 0.691, 95% CI 0.612–0.769 vs 0.526, p<0.001). Both scores had low negative predictive value (CHOKAI 47.3%, STONE 28.1%). The application of a CHOKAI-based rule-out strategy would have resulted in 29 missed CT-confirmed calculi (52.7%), suggesting that neither CHOKAI nor STONE is sufficiently reliable to replace CT-KUB for exclusion of ureteric calculi.

Conclusion: CHOKAI demonstrated significantly superior diagnostic discrimination compared with STONE for ureteric calculi in this South Asian population. However, its limited sensitivity indicates that neither score can safely replace CT-KUB for exclusion of calculi. CHOKAI may serve as a useful non-invasive adjunct for risk stratification and CT triage, particularly in resource-limited settings, pending prospective validation.

KEYWORDS: CHOKAI score, STONE score, ureteric calculi, CT-KUB, diagnostic accuracy, renal colic, Pakistan

INTRODUCTION

Regionally, kidney-stone / Ureteric stone disease is also highly prevalent. Pakistani and South-Asian data suggest that the prevalence of ureteric stone lies between 1–5 % in the general population, with one CT-based series reporting 2.8 % incidentally detected, asymptomatic stones in adults undergoing abdominal CT for other reasons (1, 2). A review of Asian data estimated that overall urolithiasis prevalence is highest in in South Asia, followed by West Asia and Southeast Asia, compared with regions of East and North Asia (3). Within Pakistan, a survey cited by Bashir et al. (2022) reported an approximate 16 % prevalence in some local populations, emphasizing the country's position within the "stone belt"(4). South Asia also carries a disproportionate paediatric burden as Global Burden of Disease estimates for 2021 indicate 1.37 million paediatric stone cases in the region, the highest worldwide (5). Regional studies have also highlighted the association between urinary stone disease and metabolic abnormalities. In a hospital-based study of 118 patients with urinary tract stones, metabolic syndrome was present in 29.7% of patients. Among those with metabolic syndrome, 54.3% had reduced HDL cholesterol, while 60.0% had elevated triglyceride levels. These findings suggest a potential link between metabolic abnormalities and urinary stone disease.(6)

The STONE score was derived and prospectively validated in a multicenter cohort of 2014 from the BMJ (7) and is based on five easily-available variables (sex, pain duration, race, nausea/vomiting, microscopic hematuria) which are combined into a 0-13 score that categorizes patients into low-risk and high-risk groups. Later research criticized the use of race, published studies touted the problem with potential spectrum bias and modest specificity, leading to a number of modifications. Fukuhara et al. suggested the CHOKAI score in 2017, which included the addition of hydronephrosis (as seen by ultrasound) and age and prior stone history, and the

removal of race. The internal validation of the model was impressive with AUC of 0.97, a sensitivity of 93 % and specificity of 90 % at cut-off ≥6 (4). The external validation of CHOKAI was conducted in Japanese, Korean, and Turkish cohorts and consistently showed better diagnostic accuracy than STONE, with AUCs of 0.92–0.95. A second series of 219 patients showed that the strength of CHOKAI was still present at a cut-off > 6 with 84 % sensitivity, 96.7 % specificity and LR+ 25.2 (8) .

Although the results have been promising, there are limited head to head comparisons between CHOKAI and STONE, and no data available for South Asia. Both scores need to be validated locally as per the difference in stone epidemiology and the fact that CT-KUB is routinely used in the emergency departments in Pakistan. A successful method, but with high radiation doses, cost and load on the scanner. The race variable of the STONE score does not apply to South Asian populations and CHOKAI makes use of ultrasound findings and stone history. Clinicians are not sure which test to use for safe imaging decision-making without local validation, which is contributing to overuse of CT. This study thus is designed to compare diagnostic accuracy of CHOKAI and STONE to non-contrast CT-KUB as a reference standard, calculate their sensitivity, specificity, positive predictive value and negative predictive value and evaluate the effectiveness of CHOKAI to help reduce unnecessary CT-KUB imaging in an emergency department in Pakistan.

MATERIALS AND METHODS

This is a cross-sectional study that was carried out in Department of Urology in Ziauddin Hospital, Karachi with the approval of the institutional Ethical Review Committee (12540526SAURO). Adult patients (aged 18 years and older) who present with acute renal or ureteric colic that is suspicious of the presence of a ureteric calculus were consecutively enrolled, with informed written consent, following the identification of a suspected calculus. Patients under 18 years old and with contraindications to CT scanning (e.g., pregnancy) were not included in the study. Patient confidentiality was ensured by assigning different numbers to each participant in the study and only the research team members had access to the data collected. The research was carried out based on the ethical principles of the Declaration of Helsinki.

Eligible patients fulfilling the inclusion criteria were recruited using a consecutive sampling technique until the required sample size was achieved. The sample size was calculated using Buderer's formula for diagnostic accuracy studies by incorporating the expected sensitivity and specificity of the index test, disease prevalence, confidence level, and desired precision (9). The expected sensitivity (93%) and specificity (90%) of the CHOKAI score were obtained from the original internal validation study conducted by Fukuhara et al. (2020), which evaluated the score for the prediction of ureteric calculi (10). A 95% confidence level and an absolute precision of 7% were assumed, while the prevalence of ureteral stones was taken as 67.74% based on the same study (11). Using the larger of the sensitivity-based and specificity-based estimates, the minimum required sample size was calculated to be 219 participants. As the present study follows a cross-sectional design, no additional allowance for attrition or dropout was included. The sample size calculation was further verified using an online diagnostic test sample size calculator based on Buderer's method (12).

Patients presenting with acute ureteric colic, defined as severe flank pain originating from the costovertebral angle and radiating toward the groin or testicle, and fulfilling the eligibility criteria were invited to participate in the study. Written informed consent was obtained from all participants before enrolment and before the use of their clinical information for research purposes. Following enrolment, demographic information was recorded for each participant. Height was measured in meters using a wall-mounted measuring scale, while body weight was measured to the nearest kilogram using a weighing machine. Body mass index (BMI) was subsequently calculated, and participants were categorized as obese if the BMI was greater than 30 kg/m² and non-obese if the BMI was 30 kg/m² or below.

Clinical history and physical examination findings were assessed by the researcher to calculate both the CHOKAI and STONE scores. A CHOKAI score of 10 or greater was considered positive, whereas a score below 10 was considered negative. Similarly, a STONE score of 10 or greater was classified as positive, while a score below 10 was considered negative. All enrolled participants subsequently underwent non-contrast CT-KUB examination, which served as the reference standard. The CT-KUB images were interpreted by a radiologist with more than ten years of experience, and the presence of a hyperdense ureteric calculus was considered a positive finding according to the predefined operational criteria. For patients with a positive CT-KUB, the maximum stone diameter, measured in millimeters, was documented from the radiology report. Stones were subsequently categorized into two groups, namely less than 3 mm and 3 mm or greater, for subgroup analysis. In addition, transabdominal ultrasonography was performed to assess the presence or absence of hydronephrosis. Ultrasonographic findings were recorded as positive or negative, with positive findings defined as the presence of hydronephrosis on transabdominal ultrasonography. Quantitative variables including age, height, weight, body mass index, CHOKAI score, STONE score, and maximum stone diameter, along with qualitative variables including gender, residential status, side of ureter involvement, number of stones, family history of renal stones, ultrasonography findings, CT-KUB findings, CHOKAI score category, STONE score category, and stone size category, were recorded in the study proforma attached as Annexure.

The collected data were entered and analyzed using SPSS version 27. Quantitative variables, including age, height, weight, body mass index, CHOKAI score, STONE score, and maximum stone diameter, were summarized as mean ± standard deviation if normally distributed or as median and interquartile range if non-

normally distributed. Qualitative variables, including gender, residential status, family history of renal stones, side of ureter involvement, ultrasonography findings, CT-KUB findings, CHOKAI score category, STONE score category, and stone size category, were presented as frequencies and percentages. The diagnostic performance of the CHOKAI and STONE scores was evaluated using non-contrast CT-KUB as the reference standard by calculating sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, and negative likelihood ratio with corresponding 95% confidence level. Receiver operating characteristic (ROC) curves were constructed for both scoring systems, and the areas under the curve (AUCs) were compared. As DeLong's test and Net Reclassification Improvement (NRI) cannot be performed in SPSS, these analyses were conducted using R software. Subgroup analyses were performed according to stone size (<3 mm versus ≥3 mm) and body mass index (>30 kg/m² versus ≤30 kg/m²). A p-value of less than 0.05 was considered statistically significant. Records with incomplete proformas or missing essential diagnostic variables were excluded from the final analysis.

RESULTS

The median age of the study population (n = 231) was 38.0 (IQR=30-52) years. 165 (71.4%) were males, residents of urban areas (n=173, 74.9%). The median body mass index (BMI) was 26.2 kg/m² (IQR 23.6–29.1), and 40 patients (17.3%) had a BMI exceeding 30 kg/m². A family history of renal stones was present in 47 patients (20.3%), and stones were slightly more frequent on the left side (113, 49.1%) than the right (106, 46.1%), with bilateral involvement in 11 patients (4.8%). Non-contrast CT-KUB, the reference standard, confirmed ureteric calculi in 170 patients (73.6%), while 61 patients (26.4%) had negative studies. The median CHOKAI score was 12.0 (IQR 10.0–12.0), and 176 patients (76.2%) were classified as high risk (≥10), the median STONE score was 6.0 (IQR 4.0–8.0; range 3–13), with only 35 patients (15.2%) classified as high risk (≥10). Among the 170 CT-positive patients, the median maximum stone diameter was 4.6 mm (IQR 4.0–5.8), and most stones (149, 87.6%) measured ≥3 mm. Baseline, anthropometric, clinical, and scoring characteristics are summarized in Table 1.

Table 1. Baseline and Clinical Characteristics of the Study Population

Variable	Descriptive statistics (n = 231)
Demographic characteristics	
Age (years), median (IQR)	38.0 (30.0–52.0)
Male, n (%)	165 (71.4)
Female, n (%)	66 (28.6)
Urban residence, n (%)	173 (74.9)
Rural residence, n (%)	58 (25.1)
Anthropometric characteristics	
Height (cm), median (IQR)	165.0 (158.4–170.7)
Weight (kg), median (IQR)	70.0 (63.0–80.0)
BMI (kg/m²), median (IQR)	26.2 (23.6–29.1)
BMI ≤30 kg/m², n (%)	191 (82.7)
BMI >30 kg/m², n (%)	40 (17.3)
Clinical characteristics	
Family history of renal stones, n (%)	47 (20.3)
Left-sided stone, n (%)	113 (49.1)
Right-sided stone, n (%)	106 (46.1)
Bilateral involvement, n (%)	11 (4.8)
Reference standard (non-contrast CT-KUB)	
CT-KUB positive, n (%)	170 (73.6)
CT-KUB negative, n (%)	61 (26.4)
Scoring systems	
CHOKAI score, median (IQR)	12.0 (10.0–12.0)
CHOKAI high risk (≥10), n (%)	176 (76.2)
CHOKAI <10 (not high risk), n (%)	55 (23.8)
STONE score, median (IQR)	6.0 (4.0–8.0)
STONE high risk (≥10), n (%)	35 (15.2)
STONE moderate risk (6–9), n (%)	87 (37.7)
STONE low risk (0–5), n (%)	109 (47.2)
Stone characteristics (CT-KUB–positive patients only; n = 170)	
Maximum stone diameter (mm), median (IQR)	4.6 (4.0–5.8)
Range (min–max)	2.0–8.0
Stone <3 mm, n (%)	21 (12.4)
Stone ≥3 mm, n (%)	149 (87.6)

IQR, interquartile range; BMI, body mass index; CT-KUB, computed tomography of the kidneys, ureters, and bladder.

At the pre-specified threshold of ≥10, the CHOKAI score demonstrated markedly higher sensitivity than the STONE score (82.9%, 95% CI 76.6-87.9% versus 17.1%, 95% CI 12.1-23.4%), correctly identifying 141 of the 170 CT-confirmed calculi compared with only 29 identified by the STONE score. Conversely, the STONE score achieved higher specificity (90.2%, 95% CI 80.2-95.4% versus 42.6%, 95% CI 31.0-55.1% for CHOKAI). Positive predictive values were similar between the two scores (CHOKAI 80.1%, 95% CI 73.6-85.3%; STONE 82.9%, 95% CI 67.3–91.9%), whereas negative predictive values were low for both (CHOKAI 47.3%, 95% CI 34.7–60.2%; STONE 28.1%, 95% CI 22.2–34.7%), indicating that neither score reliably excluded ureteric calculi. Overall diagnostic accuracy was significantly higher for CHOKAI (72.3%, 95% CI 66.2-77.7%) than for STONE (36.4%, 95% CI 30.4-42.7%). Positive and negative likelihood ratios were 1.45 (95% CI 1.15-1.81) and 0.40 (95% CI 0.26-0.62) for CHOKAI, and 1.73 (95% CI 0.76–3.97) and 0.92 (95% CI 0.83–1.02) for STONE, respectively. McNemar's test confirmed a statistically significant difference between the paired classifications of the two scores ($\chi 2 = 135.172$, $p < 0.001$). 143 patients were classified as positive by CHOKAI but negative by STONE, compared with only two classified in the opposite direction. Diagnostic Performance of CHOKAI and STONE Scores are summarized in Table 2.

Table 2. Diagnostic Performance of CHOKAI and STONE Scores against Non-Contrast CT-KUB

Diagnostic measure	CHOKAI ≥10	STONE ≥10
True positives	141	29
True negatives	26	55
False positives	35	6
False negatives	29	141
Diagnostic performance parameters (95% CI)		
Sensitivity, %	82.9 (76.6-87.9)	17.1 (12.1-23.4)
Specificity, %	42.6 (31.0-55.1)	90.2 (80.2-95.4)
Positive predictive value, %	80.1 (73.6-85.3)	82.9 (67.3-91.9)
Negative predictive value, %	47.3 (34.7-60.2)	28.1 (22.2-34.7)
Diagnostic accuracy, %	72.3 (66.2-77.7)	36.4 (30.4-42.7)
Positive likelihood ratio	1.45 (1.15-1.81)	1.73 (0.76-3.97)
Negative likelihood ratio	0.40 (0.26-0.62)	0.92 (0.83-1.02)
McNemar's test (paired comparison)		
Discordant pairs, CHOKAI+/STONE−	143	-
Discordant pairs, CHOKAI−/STONE+	-	2
χ^2 (continuity-corrected)	135.172	
p-value	<0.001	

CI, confidence interval (Wilson score method). McNemar's test applied with continuity correction to paired binary classifications.

The CHOKAI score achieved a significantly higher AUC (0.691, 95% CI: 0.612–0.769) than the STONE score (0.526, 95% CI: 0.443–0.609). The CHOKAI AUC was significantly greater than the chance level of 0.50 ($p < 0.001$), indicating moderate discriminatory ability, whereas the STONE AUC did not differ significantly from chance ($p = 0.268$). Direct comparison using DeLong's test for correlated ROC curves confirmed a statistically significant difference between the two AUCs (AUC difference: 0.165, 95% CI: 0.072–0.257; $Z = 3.480$, $p < 0.001$). The Youden Index identified exploratory, data-derived optimal thresholds of ≥12 for CHOKAI (sensitivity: 74.1%; specificity: 68.9%) and ≥9 for STONE (sensitivity: 20.6%; specificity: 86.9%). These exploratory thresholds were reported separately from the pre-specified threshold analysis. Full ROC parameters are summarized in Table 3**.**

Table 3. ROC Analysis and Comparison of CHOKAI and STONE Scores

ROC parameter	CHOKAI score	STONE score
Area under the ROC curve (AUC)		
AUC	0.691	0.526
Standard error (DeLong)	0.040	0.042
95% CI	0.612–0.769	0.443–0.609
p-value (AUC vs 0.50)	<0.001	0.268
Data-derived optimal threshold (Youden index)		
Optimal threshold	≥12	≥9
Sensitivity at optimal threshold	74.1%	20.6%
Specificity at optimal threshold	68.9%	86.9%
Pre-specified threshold performance (≥10)		
Sensitivity	82.9%	17.1%

Specificity	42.6%	90.2%
DeLong's test for correlated ROC curves		
AUC difference (CHOKAI − STONE)	0.165	
Standard error of difference	0.047	
95% CI of difference	0.072–0.257	
Z statistic	3.480	
p-value	<0.001	

AUC, area under the ROC curve; CI, confidence interval (DeLong method).

The ROC curves for CHOKAI and STONE scores are presented in **Figure 1**, illustrating their discriminatory performance against non-contrast CT-KUB.

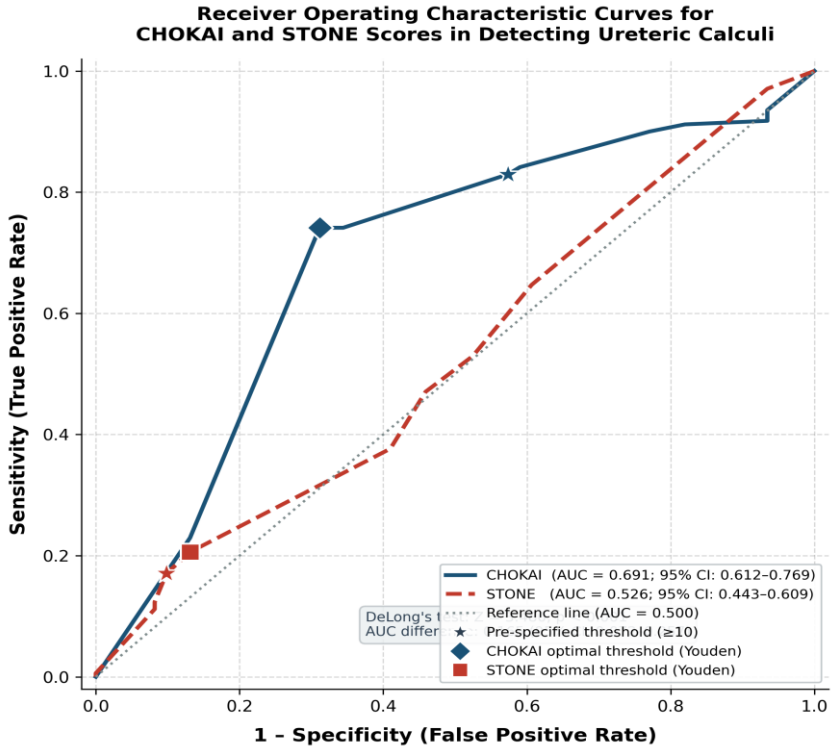


Figure 1. Receiver Operating Characteristic Curves for CHOKAI and STONE Scores in Detecting Ureteric Calculi

ROC curves comparing CHOKAI (blue solid line; AUC = 0.691) and STONE (red dashed line; AUC = 0.526) scores against non-contrast CT-KUB. The grey dotted diagonal represents chance (AUC = 0.500). Stars denote performance at the pre-specified threshold of ≥10; diamond and square markers denote the data-derived optimal thresholds (Youden index) for CHOKAI and STONE, respectively. DeLong's test: Z = 3.480, p < 0.001.

At the pre-specified threshold of ≥10, the diagnostic performance of the CHOKAI and STONE scores is compared in **Figure 2**.

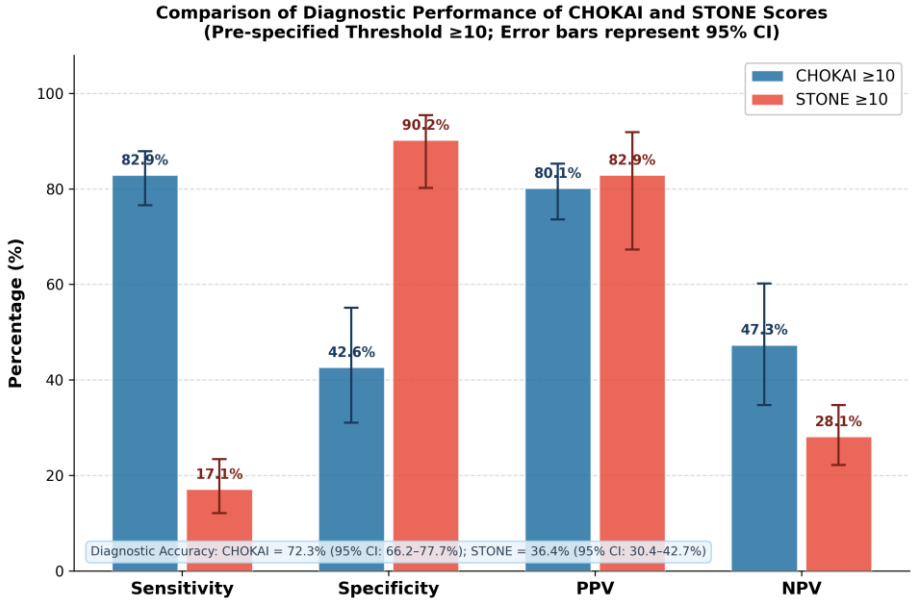


Figure 2. Comparison of Diagnostic Performance of CHOKAI and STONE Scores

Grouped bar chart comparing sensitivity, specificity, positive predictive value, and negative predictive value of CHOKAI ≥10 (blue) and STONE ≥10 (red) against non-contrast CT-KUB (n = 231). Error bars represent 95% Wilson score confidence intervals.

The exploratory analysis examined whether a CHOKAI score <10 could identify patients in whom CT-KUB might be withheld. Of the 231 patients, 55 (23.8%) had a CHOKAI score <10, among these, 29 (52.7%) nevertheless had CT-confirmed ureteric calculi, corresponding to a false-negative rate of 52.7% and consistent with the low negative predictive value of the CHOKAI ≥10 strategy (47.3%, 95% CI 34.7-60.2%). Although this approach could theoretically have avoided CT-KUB in 26 of 231 patients (11.3%), the associated risk of missing 29 confirmed cases indicates that a CHOKAI-based rule-out strategy is not clinically safe in this population. Net reclassification improvement was not calculated, as the CHOKAI and STONE scores are ordinal clinical instruments rather than probabilistic regression models, precluding valid estimation of individual-level reclassification; comparative performance at the shared threshold of ≥10 was instead assessed using McNemar's test, as reported above.

The subgroup of patients with stones <3 mm (n = 82), CHOKAI ≥10 retained a sensitivity of 90.5% (95% CI 71.1-97.3%), whereas STONE ≥10 failed to detect any of the 21 small stones (sensitivity 0.0%, 95% CI 0.0-15.5%), the corresponding AUC was 0.671 for CHOKAI and 0.438 for STONE, the latter falling below the level expected by chance. In the subgroup with stones ≥3 mm (n = 210), sensitivity and specificity for both scores were consistent with the overall cohort (CHOKAI: sensitivity 81.9%, specificity 42.6%, AUC 0.694; STONE: sensitivity 19.5%, specificity 90.2%, AUC 0.539). These findings were similar across BMI strata, among patients with BMI ≤30 kg/m² (n = 191), CHOKAI sensitivity and specificity were 83.0% and 44.0% (AUC 0.701), compared with 17.0% and 90.0% for STONE (AUC 0.539); among patients with BMI >30 kg/m² (n = 40), CHOKAI sensitivity and specificity were 82.8% and 36.4% (AUC 0.647), and STONE values were 17.2% and 90.9% (AUC 0.473), although the small number of CT-negative patients (n = 11) in this subgroup limits the precision of the specificity and NPV estimates as presented in Table 4.

Table 4. Diagnostic Performance of CHOKAI and STONE Scores According to Stone-Size and BMI Subgroups

Subgroup	Score	Sensitivity, % (95% CI)	Specificity, % (95% CI)	AUC
Stone <3 mm (n = 82)	CHOKAI ≥10	90.5 (71.1–97.3)	42.6 (31.0–55.1)	0.671
	STONE ≥10	0.0 (0.0–15.5)	90.2 (80.2–95.4)	0.438†
Stone ≥3 mm (n = 210)	CHOKAI ≥10	81.9 (74.9–87.2)	42.6 (31.0–55.1)	0.694
	STONE ≥10	19.5 (13.9–26.6)	90.2 (80.2–95.4)	0.539
BMI ≤30 kg/m² (n = 191)	CHOKAI ≥10	83.0 (75.9–88.3)	44.0 (31.2–57.7)	0.701
	STONE ≥10	17.0 (11.7–24.1)	90.0 (78.6–95.7)	0.539
BMI >30 kg/m² (n = 40)	CHOKAI ≥10	82.8 (65.5–92.4)	36.4 (15.2–64.6)	0.647
	STONE ≥10	17.2 (7.6–34.5)	90.9 (62.3–98.4)	0.473‡

AUC, area under the ROC curve; BMI, body mass index. †AUC below the chance level (0.50) for STONE in the <3 mm subgroup. ‡Small cell sizes in the BMI >30 kg/m² subgroup (CT-negative n = 11) limit the reliability of specificity estimates; all subgroup analyses are exploratory.

The CHOKAI score was significantly higher in patients with CT-confirmed calculi (median 12.0, IQR 10.0-12.0) than in those without ($p < 0.001$), whereas the STONE score did not differ significantly between CT-positive and CT-negative patients ($p = 0.536$). No significant correlation was observed between the CHOKAI score and maximum stone diameter (Spearman $\rho = 0.037$, $p = 0.631$), whereas the STONE score showed a statistically significant but weak positive correlation with stone diameter (Spearman $\rho = 0.257$, $p < 0.001$).

DISCUSSION

This study demonstrated that both CHOKAI and STONE scores were significantly higher among patients with CT-confirmed ureteric calculi, supporting their ability to discriminate between stone-positive and stone-negative patients. However, CHOKAI consistently outperformed STONE across sensitivity, specificity, predictive values, likelihood ratios, and overall discrimination. This finding aligns with previous studies. (11, 13) The differences in patient selection, disease prevalence, clinical setting, ultrasound assessment, and threshold selection may partly explain this variation. These findings collectively support the reproducibility of the relative advantage of CHOKAI over STONE across different populations.

The superiority of CHOKAI may partly relate to differences in the variables incorporated into the prediction rules. The STONE score was developed from five clinical variables including sex, duration of pain, race, nausea/vomiting, and hematuria. (14) Whereas CHOKAI combines clinical characteristics with ultrasonographic assessment, including hydronephrosis. The incorporation of an objective imaging feature may provide additional information regarding urinary tract obstruction that is not captured by symptoms alone. In our cohort, ultrasonographic findings were strongly associated with CT-confirmed calculi, further supporting the potential contribution of ultrasound-derived information to CHOKAI performance. Moreover, subsequent external validation studies demonstrated more modest discrimination of the STONE score, with an AUC of approximately 0.78, and indicated that STONE could not reliably identify a sufficiently low-risk group in whom CT could safely be deferred. (15, 16) Thus, while STONE indicates useful diagnostic discrimination, the consistently better performance of CHOKAI suggests an advantage to incorporating additional objective findings.

The stone-size subgroup analysis provides further comparative evidence**.** Both prediction rules showed reduced sensitivity for calculi <3 mm, but CHOKAI remained more sensitive than STONE, consistent with previous studies. (11, 17, 18) This finding is clinically plausible because very small calculi may produce less prominent obstruction and therefore fewer detectable clinical or ultrasonographic features. Conversely, sensitivity increased substantially for stones ≥3 mm, particularly with CHOKAI. A previous prospective study similarly reported a positive correlation between CHOKAI score and stone size, supporting the possibility that CHOKAI may perform differently according to stone burden. (17)

An additional strength of our findings is the comparative assessment across BMI categories. CHOKAI maintained sensitivity comparable to the overall cohort among obese participants, although its specificity was lower than in the cohort as a whole, whereas STONE retained relatively high specificity in this subgroup but continued to show poor sensitivity and a lower negative predictive value than CHOKAI.

The potential reduction in CT utilization is clinically relevant because CT remains the reference standard for diagnosing ureteric calculi but involves ionizing radiation and additional healthcare resources. Previous research has therefore explored prediction rules and ultrasonography as strategies for reducing unnecessary CT imaging. (19) Studies evaluating ultrasound in combination with clinical risk stratification have demonstrated that ultrasonography can meaningfully modify the probability of ureteral stones, although CT generally retains greater diagnostic accuracy. (20, 21)

The study limitations include the single-center study design and the high prevalence of CT-confirmed ureteric calculi, which may limit the generalizability of findings. Retrospective extraction of demographic data from free-text fields may have introduced minor misclassification, and ultrasonographic findings were unavailable. Subgroup analyses were limited by small cell sizes, particularly for stones <3 mm and patients with BMI >30 kg/m², resulting in wider confidence intervals. Net reclassification improvement was not assessed because CHOKAI and STONE are ordinal, non-probabilistic scores. The exploratory CT-KUB reduction analysis should not be interpreted as supporting the routine withholding of CT-KUB, and larger prospective studies are required to validate these findings before clinical implementation. Future studies should prospectively validate CHOKAI-guided imaging strategies across diverse, independent populations to determine their safety and clinical utility.

CONCLUSION

CHOKAI demonstrated relatively superior diagnostic performance compared with STONE for predicting CT-confirmed ureteric calculi, with higher sensitivity, specificity, predictive values, and overall discriminations. The incorporation of ultrasonographic findings, particularly hydronephrosis, may provide additional diagnostic information and improve risk stratification compared with clinical characteristics alone. Although exploratory findings suggest potential for reducing unnecessary CT-KUB use, larger prospective validation studies are required before CHOKAI-guided imaging strategies can be recommended for routine clinical implementation.

REFERENCES

1. Stamatelou K, Goldfarb DS. Epidemiology of Kidney Stones. Healthcare (Basel). 2023;11(3).
2. Saeed S, Ullah A, Ahmad J, Hamid S. The Prevalence of Incidentally Detected Urolithiasis in Subjects Undergoing Computerized Tomography. Cureus. 2020;12(9):e10374.
3. Akram M, Somani B. Epidemiology and Management of Kidney Stone Disease - Current Insights. Res Rep Urol. 2025;17:449-59.
4. Bashir A, Zuberi SK, Musharraf B, Khan H, Ather MH. Perception of Dietary Influences on Renal Stone Formation Among the General Population. Cureus. 2022;14(6):e26024.
5. Chen S, Ma X, Guo L, Wang S, Wu J, Wu L, et al. The global, regional, and national burden of pediatric stone disease: 1990–2021 and projections for the next two decades. Frontiers in Pediatrics. 2025;Volume 13 - 2025.
6. Ali R, Sigdel M. Metabolic Syndrome in patients with Urinary Tract Stone Disease: A Hospital based Single Center Study. Journal of Institute of Medicine Nepal. 2023;45:1-6.
7. Hassouba ON, Abdullah Omar AA, Awan M, Ahmad S, Taha M, Venkatachalapathi S, et al. STONE Score as a Triage Tool to Guide Computed Tomography of the Kidneys, Ureters, and Bladder (CT-KUB) Requests in Suspected Renal Colic: A Quality Improvement Initiative. Cureus. 2025;17(9):e92080.
8. Ok F, Durmuş E. External Validation of CHOKAI and STONE Scores for Detecting Ureter Stones in the Eastern Turkish Population. Journal of Urological Surgery. 2022;9:260-4.
9. Bujang MA, Adnan TH. Requirements for Minimum Sample Size for Sensitivity and Specificity Analysis. J Clin Diagn Res. 2016;10(10):Ye01-ye6.
10. Buderer NM. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. Acad Emerg Med. 1996;3(9):895-900.
11. Fukuhara H, Kobayashi T, Takai S, Tawara T, Kikuta M, Sugiura A, et al. External validation of the CHOKAI score for the prediction of ureteral stones: A multicenter prospective observational study. Am J Emerg Med. 2020;38(5):920-4.
12. UPMED. Sensitivity / Specificity Sample Size Calculator: UPMED Medical Consultancy (Private) Limited; 2024 [updated December 31, 2024July 9, 2026]. Available from: https://upmed.net/sensitivity-specificity-sample-size-calculator.
13. Rehman MA, Ahmed T, AlZaabi A, Almarzooqi MA, Husain MM, AlNuaimi R, et al. Comparing CHOKAI and STONE Scores in Predicting Ureteral Stones in The Emergency Department: A Prospective Multicenter Diagnostic Study. The Journal of emergency medicine. 2025;78:302-12.
14. Moore CL, Bomann S, Daniels B, Luty S, Molinaro A, Singh D, et al. Derivation and validation of a clinical prediction rule for uncomplicated ureteral stone--the STONE score: retrospective and prospective observational cohort studies. BMJ (Clinical research ed). 2014;348:g2191.
15. Safaie A, Mirzadeh M, Aliniagerdroudbari E, Babaniamansour S, Baratloo A. A clinical prediction rule for uncomplicated ureteral stone: The STONE score; a prospective observational validation cohort study. Turkish journal of emergency medicine. 2019;19(3):91-5.
16. Malik A, Mohkumuddin S, Yousaf S, Baig MAR, Afzal A. Validity of STONE Score in Clinical Prediction of Ureteral Stone Disease. Pakistan journal of medical sciences. 2020;36(7):1693-7.
17. Eraybar S, Yuksel M. The prospective evaluation of the effectiveness of scoring systems in the emergency department in cases with suspected ureteral stones: STONE? CHOKAI? Am J Emerg Med. 2021;49:94-9.
18. Fukuhara H, Ichiyanagi O, Midorikawa S, Kakizaki H, Kaneko H, Tsuchiya N. Internal validation of a scoring system to evaluate the probability of ureteral stones: The CHOKAI score. The American Journal of Emergency Medicine. 2017;35(12):1859-66.
19. Daniels B, Gross CP, Molinaro A, Singh D, Luty S, Jessey R, 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. Annals of emergency medicine. 2016;67(4):439-48.
20. Cochon L, Smith J, Baez AA. Bayesian comparative assessment of diagnostic accuracy of low-dose CT scan and ultrasonography in the diagnosis of urolithiasis after the application of the STONE score. Emergency radiology. 2017;24(2):177-82.
21. Wong C, Teitge B, Ross M, Young P, Robertson HL, Lang E. The Accuracy and Prognostic Value of Point-of-care Ultrasound for Nephrolithiasis in the Emergency Department: A Systematic Review and Meta-analysis. Acad Emerg Med. 2018;25(6):684-98.