Received: 21 November 2023 | Revised: 11 September 2024 | Accepted: 24 September 2024

DOI: 10.1002/emp2.13324

JACEP OPEN
JOURNAL OF THE AMERICAN COLLEGE OF EMERGENCY PHYSICIANS OPEN
A GLOBAL JOURNAL OF EMERGENCY MEDICINE
WILEY

ORIGINAL ARTICLE

Diversity, Equity and Inclusion

Construction and performance of a clinical prediction rule for ureteral stone without the use of race or ethnicity: A new STONE score

Christopher L. Moore MD[1] | Cary P. Gross MD[2] | Louis Hart MD[3] | Annette M. Molinaro PhD[4] | Deborah Rhodes MD[2] | Dinesh Singh MD[5] | Cristiana Baloescu MD[1]

[1]Department of Emergency Medicine, Yale School of Medicine, New Haven, Connecticut, USA

[2]Department of Internal Medicine, Yale School of Medicine, New Haven, Connecticut, USA

[3]Department of Pediatrics, Yale School of Medicine, New Haven, Connecticut, USA

[4]Department of Neurological Surgery, University of California San Francisco, San Francisco, California, USA

[5]Department of Urology, Yale School of Medicine, New Haven, Connecticut, USA

Correspondence
Christopher L. Moore, MD, Department of Emergency Medicine, Yale School of Medicine, New Haven, CT, USA.
Email: chris.moore@yale.edu

Funding information
Agency for Healthcare Research and Quality, Grant/Award Number: 5R01HS018322-03

Abstract

Objectives: The original STONE score was designed to predict the presence of uncomplicated renal colic and the corresponding absence of alternate serious etiologies. It was retrospectively derived and prospectively validated and resulted in five variables: Sex (male gender), Timing (acute onset of pain), "Origin" (non-Black race), Nausea/vomiting (present), and Erythrocytes (microscopic hematuria). With recent increased awareness of the potential adverse impacts of including race (a socially constructed identity) in clinical prediction rules, we sought to determine if a revised STONE score without race could be constructed with similar diagnostic accuracy.

Methods: We used data from the original STONE score that utilized retrospective data on patients with confirmed kidney stone by computed tomography (CT) to derive a clinical prediction rule as well as prospective data to validate the score. These data were used to construct a revised STONE score after removing race as a variable. We performed univariate and multivariable logistic regression and compared the old and new STONE scores (including multivariable, integral, and three-level risk) using the area under the receiver operating characteristic curve (AUC) and misclassification rates.

Results: After the elimination of race, multivariable logistic regression revealed that gross hematuria was the next strongest feasible variable for the prediction of ureteral stone. This was incorporated into a revised STONE score by substituting "obvious hematuria" for "origin" (formerly race). The revised STONE score had similar predictive accuracy to the original STONE score: AUC 0.85 versus 0.86 (95% confidence interval [CI]: 0.82–0.87 and 0.79–0.93); misclassification rates were also unchanged, 0.23 versus 0.23 (95% CI: 0.20–0.25 and 0.20–0.25).

Supervising Editor: Yiju Teresa Liu, MD and Henry Wang, MD, MS

Conclusions: We modified the STONE score to remove race and include "obvious hematuria" without losing clinical accuracy. Considering the potential adverse effects of propagating racial bias in clinical algorithms, we recommend using the revised STONE score. Future research could investigate the potential contributions of social drivers of health (SDOH) to the diagnosis of kidney stone.

1 | INTRODUCTION

1.1 | Background

Although the debate about the appropriateness of incorporating race and ethnicity into medicine is not new, the harms of perpetuating a conceptualization of race as a biologic category have become more widely appreciated.[1] Using race in clinical medicine requires clinicians to interpret a patient's "race" and can lead to disparate care based on physical characteristics (ie, skin color, hair texture, eye shape) or personally held socio-cultural identity. Accordingly, the medical community has re-examined the use of race in evaluating and treating patients. These concerns have paralleled genomic research highlighting the imperfect relationship between race and genetics, further emphasizing that race has been socially constructed rather than biologically determined.[2]

1.2 | Importance

One domain of medicine that has received considerable attention is the use of race in clinical algorithms and prediction tools. A prominent example is kidney disease. There is a burgeoning recognition that including race in equations that estimate kidney function (eGFR) can perpetuate the fallacy of race as a biologic construct, leading to unequal care recommendations based on patient race. In 2021, a joint National Kidney Foundation (NKF) and American Society of Nephrology (ASN) task force found that a race-free CKD-EPI eGFR equation performed similarly to the previous equations that included race.[3]

This pathway toward removing race from the eGFR equation raises the question about other clinical algorithms.[4] One recent study found that removing race from an equation that predicts cancer recurrence led to a decrease in accuracy and mischaracterizing risk for racial and ethnic minoritized patients.[5] This study did not look at the potential contributions of social drivers of health (SDOH) that could be associated with race. Assessing the accuracy and clinical impact of removing race from validated algorithms is an important consideration.

Several collaborators in our group developed the STONE score in 2014 to help risk stratify patients presenting to the emergency department with back pain according to the likelihood of a symptomatic ureteral stone instead of an alternate, serious cause of symptoms.[6] The objective was to avoid unnecessary imaging and radiation in patients with a high likelihood of uncomplicated renal colic and, conversely, to identify patients without ureteral stone who had a higher likelihood of an alternate diagnosis that would suggest appropriate computed tomography (CT) imaging.

Race was a significant predictor in both univariate and multivariable analysis of our original data, and prior studies had suggested a relationship between race and risk of developing kidney stones. However, even in the early 1990s, these differences were considered potentially related to bias in access to medical care as no causal biomedical mechanism had been identified.[7–10] It is possible, and even likely, that SDOH that are associated with race could impact access to and decisions to seek care and affect the presentation to the emergency department setting, making it appear that race is a causative factor.

Although not the investigators' intention, incorporating race as a variable into the clinical algorithm has the potential to propagate the harms of race-based medicine (ie, perpetuating the fallacy of race as biologically deterministic, and fostering clinical racial discrimination). Further, an external validation of the STONE score found that while four out of five predictor variables were validated, non-Black race (the "O" in the STONE score) was not significantly associated with ureteral stone.[11] The variable prevalence of race in different geographic regions also challenges the utility of a STONE score that includes race.

1.3 | Goals of this investigation

Using data from the original derivation and validation of the STONE score, we sought to determine if an alternate score that did not include race could predict uncomplicated renal colic with similar accuracy to the original STONE score.

2 | METHODS

We utilized data from the derivation and validation of the original STONE score, and did not incorporate new data. Methodology for the original study has been previously described.[6,12] Briefly, important elements of the data and analysis are described below.

2.1 | Study design

This was a secondary analysis of previously collected retrospective and prospective data used in the initial derivation and validation of the STONE score.

2.2 | Setting

Data were collected from patients in two emergency departments associated with an academic medical center, one urban ED with over

The Bottom Line

The role of race in renal colic management is unclear. In this study the authors revised the original STONE score for predicting a symptomatic kidney stone by replacing "origin or race" (O) with "obvious or gross hematuria" as a predictor. The updated STONE score showed high discrimination (AUC 0.85) similar to the original STONE score (AUC 0.86). The updated STONE score may provide the basis for future clinical application and research.

80,000 annual adult visits, and one free-standing ED with over 20,000 annual visits.

2.3 | Selection of participants

Participants were 18 years or older and had a CT scan for suspected renal colic, with a CT diagnosis of ureteral stone used as the gold standard for diagnosis. All subjects had symptoms of ureteral stone (back or flank pain), with exclusions for signs of infection (fever or leukocytes on urine dipstick), known active malignancy, known renal disease (creatinine >1.5 mg/dL), or previous urologic procedure (including lithotripsy or ureteral stent).

2.4 | Data collection and measurements

The derivation dataset included 1040 subjects. The original sample size determination was based on a prediction of intervention for ureteral colic. It yielded a sample size of approximately 1000 patients based on the assumption that about 50% of subjects receiving a CT would have a ureteral stone, and about 20% would undergo intervention.

Before data abstraction, five physician co-investigators from different specialties (urology, emergency medicine, internal medicine) created an a priori list of factors considered potentially predictive of ureteral stone. This list was developed based on clinical experience and a literature review. It included over 50 factors from demographics, history of present illness, past medical/family/social history, physical examination, and urine testing. Two trained observers abstracted data elements blinded to the CT result, with inter-rater reliability performed on 50 randomly selected records, and any elements with kappa below 0.6 excluded from consideration in the model.

Symptomatic kidney stones were defined as ureteral stones identified between the kidney and bladder. Stones found solely in the parenchyma of the kidney were not considered symptomatic. In addition to symptomatic kidney stones, we attempted to identify CT results that revealed: "acutely important alternate findings" (AIAFs). These findings would modify care or require intervention, such as appendicitis, diverticulitis, abdominal aortic aneurysm, and so forth.

The validation dataset consisted of prospectively collected data from 491 consecutive patients who underwent a CT scan for kidney stone, with data collected prior to CT results being known. Enrollment occurred on defined shifts, including overnights, weekends, and holidays, with an automatic paging system for notifications of CTs ordered for renal colic.

The Yale IRB determined the re-use of this anonymized data as exempt from IRB review.

2.5 | Construction and accuracy of the new STONE score

All variables from the retrospective dataset were considered via univariate logistic regression analysis with the estimation of prevalence and odds ratios with corresponding 95% confidence intervals. Factors with significant univariate association with ureteral stone are shown in Table 1. Prior to the performance of multivariable logistic regression, race and ethnicity were removed. Multivariable logistic regression was then performed employing forward selection and 10-fold cross-validation for model selection. The five most predictive and feasible elements were selected for inclusion in the new STONE score.

Performance of the STONE scores included estimating two measures of prediction accuracy: the misclassification rate and the area under the receiver operating characteristic curve (AUC). The best model was the one that had a low cross-validated misclassification rate and a high AUC.

Subsequently, a scoring system was built on the best model, and the risk associated with each factor was calculated via the multiple logistic regression equation using the point attribution scale based on methods from the Framingham study.[13] A weighted kappa test was used to verify the agreement between risk estimates based on the point system and those based on the multivariable logistic regression model.

In addition to estimating AUC for summarizing the model's discrimination, the Hosmer and Lemeshow test was used to test for goodness of fit and calibration. After the point system was constructed from the derivation phase but before analysis of prospective data, the research team selected three categories for risk (low, moderate, and high) in order to qualitatively stratify the probability of ureteral stone by point total in each category. Subsequently, the scoring system was applied to the prospective dataset.

3 | RESULTS

The univariate models from the derivation dataset ($n = 1040$) yielded 35 factors significantly associated with the presence of a ureteral stone (Table 1). The multivariable analysis yielded five feasible factors most significantly associated with a ureteral stone: male sex, acute onset of pain, presence of nausea or vomiting, microscopic hematuria, and gross (obvious) hematuria. Two other variables were associated with a ureteral stone with relatively high odds ratios on multivariable regression. These variables (prior ED visits and smoking status) were not

TABLE 1 Factors significantly associated with ureteral stone on univariate analysis of retrospective data.

Significant factors for presence or absence of ureteral stone (univariate analysis)				
Factor	**Baseline (OR 1.0)**	***n* (%) of total with factor present**	***n* (%) of those with factor with ureteral stone**	**OR (95% CI)**
Demographics				
Male gender	Female gender	539 (51.8%)	371 (68.8%)	3.4 (2.6–4.4)
Non-Black race	Black race	930 (89.4%)	547 (58.8%)	6.1 (3.7–9.9)
Arrival by ambulance	Arrival by other mode	156 (15.0%)	96 (61.5%)	1.4 (1.0–2.0)
History of present illness				
Any flank pain present	No flank pain present	973 (93.6%)	551 (56.6%)	3.8 (2.2–6.9)
Any back pain present	No back pain present	315 (30.3%)	134 (42.5%)	0.5 (0.4–0.6)
Symptoms lateralized	Symptoms non-lateralized	853 (82.0%)	480 (56.3%)	1.5 (1.1–2.0)
Pain onset "abrupt" or "sudden"	Pain onset gradual or unknown	630 (61.0%)	420 (66.7%)	3.5 (2.7–4.6)
Pain course "constant"	Pain course not constant	367 (35.3%)	223 (60.8%)	1.5 (1.1–1.9)
Pain with movement present	No pain with movement	222 (21.3%)	91 (41%)	0.5 (0.4–0.7)
Pain duration <6 h	Pain course 1 day to 1 week	375 (36.1%)	292 (77.9%)	5.8 (4.1–8.2)
Pain duration 6 h to 1 day	Pain course 1 day to 1 week	259 (24.9%)	137 (52.9%)	1.8 (1.3–2.6)
Pain for more than 1 week	Pain course 1 day to 1 week	113 (10.9%)	23 (20.4%)	0.4 (0.2–0.7)
Pain "severe" or 7–10 out of 10	Pain not "severe" or <7 out of 10	744 (71.5%)	445 (59.8%)	2.1 (1.6–2.8)
Radiation of pain to groin present	No radiation of pain to groin	336 (32.3%)	229 (68.2%)	2.3 (1.8–3.0)
Nausea alone present	No nausea or vomiting	311 (29.9%)	176 (56.6%)	1.9 (1.4–2.6)
Nausea with vomiting present	No nausea or vomiting	298 (28.7%)	219 (73.5%)	4.1 (3.0–5.7)
Presence of diarrhea	Absence of diarrhea	53 (5.1%)	21 (39.6%)	0.5 (0.3–0.9)
Presence of dysuria	Dysuria not present	211 (20.3%)	129 (61.1%)	1.9 (1.5–2.5)
Subjective hematuria present	Subjective hematuria not present	205 (19.7%)	139 (67.8%)	2.0 (1.5–2.8)
Past medical/family/social history				
Presence of any allergy	No allergy present	335 (32.2%)	143 (42.7%)	0.5 (0.4–0.6)
No pior ED visits documented	Prior ED visits documented	592 (56.9%)	404 (68.2%)	3.7 (2.9–4.8)
Family history of kidney stones	No family history/not mentioned	63 (6.1%)	50 (79.4%)	3.4 (1.9–6.6)
Any history of smoking present	No history of smoking	195 (18.8%)	78 (40%)	0.5 (0.4–0.7)
Prior history of kidney stones	No prior history of kidney stones	326 (31.3%)	194 (59.5%)	1.4 (1.0–1.7)
Any past surgical history present	No past surgical history	302 (29%)	141 (46.7%)	0.6 (0.5–0.8)
Taking any medication present	No medications documented	464 (44.7%)	227 (48.9%)	0.7 (0.5–0.8)
Physical examination				
Elevated systolic blood pressure, each 10 mmHg	Mean 134 ± 35 mmHg	n/a	n/a	1.2 (1.1–1.2)
Elevated diastolic blood pressure, each 10 mmHg	Mean xx ± mmHg	n/a	n/a	1.3 (1.2–1.4)
Elevated pulse, per 10 beats/min	Mean 83 ± 15 bpm	n/a	n/a	0.8 (0.8–0.9)
Right lower quadrant tenderness present	No RLQ tenderness	171 (16.4%)	107 (62.6%)	1.5 (1.1–2.1)
Right or left lower quadrant tenderness present	No right or left lower quadrant tenderness	330 (31.7%)	198 (60.0%)	1.4 (1.1–1.8)
Upper abdominal tenderness present	No upper tenderness	91 (8.8%)	38 (41.8%)	0.6 (0.4–0.9)
Presence of lumbar or back tenderness		106 (10.2%)	36 (33.0%)	0.4 (0.2–0.6)
Laboratory values				
Any urine RBCs	No RBCs	717 (68.9%)	473 (66.0%)	4.7 (3.5–6.2)
Creatinine	Each 0.1 mg/dL increase	n/a	n/a	1.3 (1.2–1.4)

Abbreviations: bpm, beats per minute; CI, confidence interval; ED, emergency department; OR, odds ratio; RBC, red blood cell; RLQ, right lower quadrant.

TABLE 2 Elements of the new STONE score with univariate and multivariate odds ratios and associated integral point values.

Retro data multivariate model					
STONE score	**KS**	**No KS**	**OR (univariable)**	**OR (multivariable)**	**Points**
Sex					
Female	305 (60.9)	196 (39.1)	–	–	0
Male	168 (31.2)	371 (68.8)	3.44 (2.67–4.45, $p < 0.001$)	4.38 (3.20–6.03, $p < 0.001$)	2
Timing					
>24 h	267 (65.8)	139 (34.2)	–	–	0
6–24 h	123 (47.5)	136 (52.5)	2.12 (1.55–2.92, $p < 0.001$)	1.88 (1.30–2.73, $p = 0.001$)	1
<6 h	83 (22.1)	292 (77.9)	6.76 (4.93–9.33, $p < 0.001$)	6.52 (4.52–9.51, $p < 0.001$)	3
Obvious hematuria					
No	407 (48.7)	428 (51.3)	–	–	0
Yes	66 (32.2)	139 (67.8)	2.00 (1.46–2.78, $p < 0.001$)	1.82 (1.23–2.72, $p = 0.003$)	1
Nausea					
None	258 (59.9)	173 (40.1)	–	–	0
Nausea alone	136 (43.7)	175 (56.3)	1.92 (1.43–2.58, $p < 0.001$)	2.07 (1.45–2.97, $p < 0.001$)	1
Nausea and vomiting	79 (26.5)	219 (73.5)	4.13 (3.01–5.72, $p < 0.001$)	4.92 (3.34–7.33, $p < 0.001$)	2
Erythrocytes					
No	228 (70.6)	95 (29.4)	–	–	0
Yes	245 (34.2)	472 (65.8)	4.62 (3.49–6.17, $p < 0.001$)	5.13 (3.65–7.28, $p < 0.001$)	2

Abbreviations: OR, odds ratio, KS, kidney stone.

included in the new model due to concerns about feasibility and validity. Previous ED visits, as ascertained from the medical record, were associated with a lower likelihood of nephrolithiasis, but were not included in the model to maximize the generalizability of the score across medical centers, given that visits to outside ED sites are frequently not captured in the medical record (this was also excluded from the original STONE score for the same reason). Smoking was associated with a lower likelihood of ureteral stone in our study. Still, an a priori decision was made not to include it in the model because this association was contrary to the literature consensus that smoking increases the likelihood of ureteral stones, and the fact that smoking history (and degree of smoking) is notoriously difficult to capture in a retrospective review, likely limiting validity.

These five factors were incorporated into a new STONE score that includes "S" for biologic sex, "T" for timing of pain, "O" for obvious or gross hematuria, "N" for nausea and vomiting, and "E" for erythrocytes or microscopic hematuria. These factors, shown with odds ratios, were assigned integral points yielding a score from 0 to 10 (Table 2) and were then stratified into low, moderate, and high risk (Table 3). Categorization into low, moderate, and high risk of ureteral stone with the likelihood of ureteral stone and AIAFs are shown in Table 4 and Figure 1.

For the derivation phase from the retrospective data, the new STONE score yielded an AUC for the multivariable model of 0.85 (95% confidence interval [CI]: 0.82–0.86) that is non-inferior to the original STONE score, which had an AUC of 0.86 (95% CI: 0.79–0.93). Receiver operator characteristic (ROC) curves for retrospective and prospective data are shown in Figures 2 and 3. Comparison of other metrics

TABLE 3 Stratified risk of ureteral stone based on 0–10 point scale in new STONE score.

Score	Points	Likelihood of KS (%)
Low	0	4
	1	7
	2	14
	3	26
Mod	4	42
	5	60
	6	76
High	7	87
	8	93
	9	96
	10	98

Abbreviation: KS, kidney stone.

of performance in AUC, misclassification was comparable with the new STONE score for both the derivation and validation phases (Table 5). Agreement between risk estimates calculated using the new STONE score and those computed using the multivariable logistic regression model had a κ (weighted) of 0.90 for the derivation phase and 0.89 for the validation phase, indicating minimal loss of accuracy with the incorporation of the integral point scale. The Hosmer–Lemeshow $\chi^2 = 6.47$ was insignificant ($p = 0.58$), indicating good discrimination and calibration.

TABLE 4 Performance of stratified STONE scores in prediction of kidney stone and AIAFs.

	Original STONE score			
	Score	**Category**	**Stone**	**AIAF**
Low	0–5	78 (15.9%)	7/78 (9.1%)	4/78 (5.1%)
Moderate	6–9	226 (46.0%)	117/226 (51.8%)	11/226 (4.9%)
High	10–13	187 (38.0%)	165/187 (88.2%)	3/187 (1.6%)
	New STONE score			
Low	0–3	154 (31.3%)	32/154 (20.8%)	11/154 (7.1%)
Moderate	4–6	202 (41.1%)	137/202 (67.8%)	4/202 (2.0%)
High	7–10	135 (27.5%)	120/135 (88.9%)	3/135 (2.2%)

Abbreviation: AIAF, acutely important alternate finding.

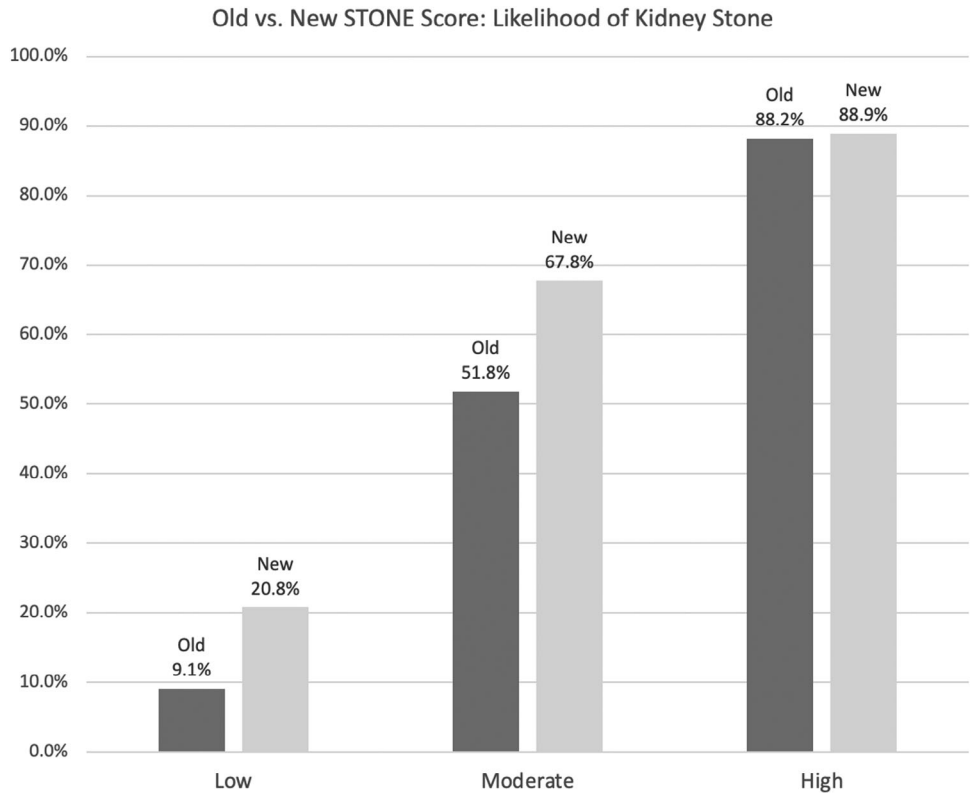

FIGURE 1 Comparison of the old versus new STONE scores stratified by likelihood into low, medium, and high likelihood of kidney stone.

4 | LIMITATIONS

This is a re-analysis of the data that were previously collected by our group. While we maintained granular data and attempted to minimize bias, the data are now 10 years old and the prior analysis and objectives were known so it is difficult to eliminate all bias without further blinded prospective data collection. While the data did come from two physically separate emergency departments, they are in the same general geographic area and may not be as generalizable to other locations. While we attempted to objectively include the most predictive values after race was excluded, we did continue to exclude possible predictors such as prior ED visits, as these could be difficult to replicate across systems.

The inclusion of both gross and microscopic hematuria does involve some degree of collinearity, as most often when there is gross hematuria there will also be microscopic hematuria. However, in the retrospective dataset, agreement (both positive or both negative) occurred in a minority of cases (43%). The presence of both increased true positives (66% with microscopic hematuria vs. 72% with both) and the absence of either decreased false negatives (29% without microscopic hematuria vs. 26% with neither). Statistically, the presence of collinearity may affect the accuracy of the regression coefficients, but should

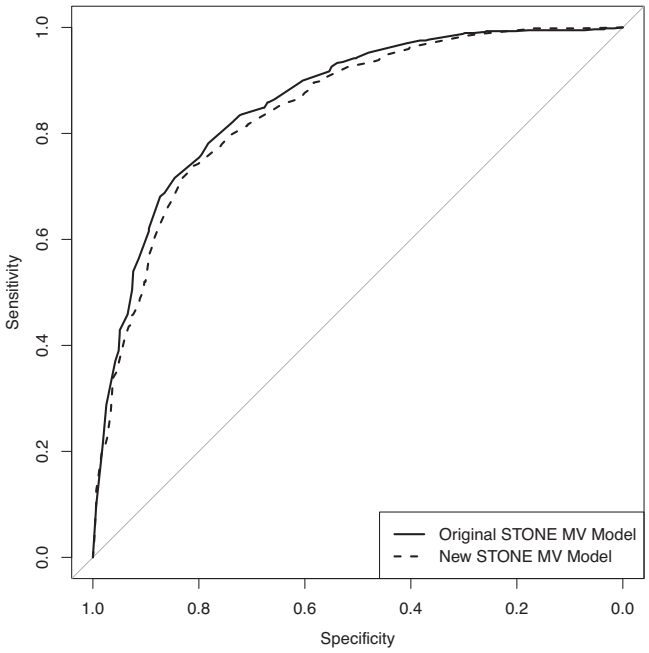

FIGURE 2 Receiver operator characteristic (ROC) curve for the retrospective data of the old and new scores.

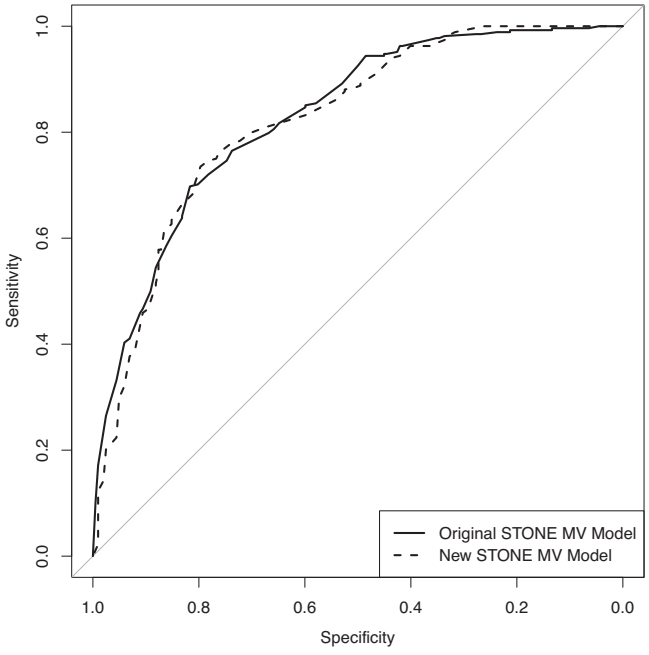

FIGURE 3 Receiver operator characteristic (ROC) curve for the prospective data of the old and new scores.

not affect the accuracy of the predictive model, and the inclusion of gross hematuria as part of the model had the best overall effect on model prediction compared to adding other variables.

It is possible that the original significance of race as a variable was due to an association of race with SDOH. Unfortunately, the initial data used to revisit the STONE score did not include important elements of SDOH such as census tract or zip code for socioeconomic status, and insurance type. A limitation of the current paper is that these may be important but unaccounted for, limiting the ability of the revised STONE score to adequately risk-stratify people with socially determined threats to their health. The original data also included gender as binary, limiting the ability to address how non-binary gender might affect the score.

5 | DISCUSSION

We set out to create a revised predictive model for uncomplicated renal colic that was reproducible, accurate, and free of racial bias. The new STONE score shows similar accuracy to the original STONE score in the ureteral colic prediction without using race as a factor. Although in the original analysis, non-Black race had the strongest association of being diagnosed with a ureteral stone in the ED (odds ratio [OR] 6.1), our newest findings show that the model can perform just as well without this variable.

Our results show that the substitution of "non-Black race" (formerly "O" for "origin"—probably an inaccurate term) with the "presence of gross hematuria" ("O" for "obvious hematuria") as a risk factor does not negatively impact the accuracy and performance of the STONE score. While the CI on the AUC of the original stone score does extend somewhat wider (suggesting that a larger sample size might provide more clarity), these CIs do overlap substantially in this fairly large study suggesting there is little difference in predictive accuracy. We thus propose the adoption and use of the new STONE score without the inclusion of race. While race appeared to be an influential variable in the initial score to predict clinical risk, given our contemporary understanding of the flaws of race-based medicine, we believe the unintended harms of using race outweigh the earlier perceived benefits.

It should be noted that the "low" category in the new STONE score does have a slightly higher point estimate for ureteral stone, but also includes more subjects and has a higher prevalence of AIAFs. While the cutoff for low/medium/high stratification is somewhat subjective (ie, low could be either 0–2 or 0–3), the intent of the "low" score is to help guide clinicians in considering that patients may have something other than kidney stone and may need additional diagnostic testing. Keeping the low score in 0–3 may capture more AIAFs.

While there is a long history of race in medicine, the issue of race in clinical prediction rules was brought to the forefront by a 2020 article in the New England Journal of Medicine entitled "Hidden in Plain Sight—Reconsidering the Use of Race Correction in Clinical Algorithms."[4] This article addressed the use of race in clinical prediction rules across disciplines, highlighting the STONE score as an example in urology. It should be noted that the NEJM incorrectly interpreted the intended use of the STONE score stating that "by assigning a lower score to Black patients, the STONE algorithm may steer clinicians away from thorough evaluation for kidney stones in Black patients." However, the intent of the STONE score is actually to identify patients in whom a kidney stone was more likely, thus decreasing the need for an unnecessary CT scan to identify alternative important

TABLE 5 Comparison of AUC and misclassification rate for old integer point score, the new multivariate logistic regression model, and the new integer point score.

Derivation (n = 1040)		
	AUC (95% CI)	**Misclassification rate (95% CI)**
Original STONE score multivariate model	0.86 (0.79–0.93)	0.23 (0.20–0.25)
Original STONE score	0.82 (0.74–0.90)	0.23 (0.20–0.25)
Kappa	0.87 (0.86–0.87)	
New multivariate model	0.85 (0.82–0.87)	0.23 (0.20–0.25)
New STONE score	0.84 (0.82–0.86)	0.23 (0.20–0.25)
Kappa	0.90 (0.89–0.90)	
New STONE score (3 levels)	0.80 (0.77–0.82)	0.25 (0.23–0.28)
Original STONE score (3 levels)	0.80 (0.78–0.82)	0.30 (0.28–0.33)
Validation (n = 491)		
	AUC (95% CI)	**Misclassification rate (95% CI)**
Original STONE score multivariate model	0.85 (0.81–0.88)	0.23 (0.19–0.27)
Original STONE score (3 levels)	0.79 (0.76–0.83)	0.30 (0.26–0.34)
New multivariate model	0.85 (0.82–0.89)	0.21 (0.17–0.24)
New STONE score (3 levels)	0.80 (0.76–0.84)	0.24 (0.20–0.28)
Kappa	0.89 (0.88–0.90)	

Abbreviations: AUC, area under the curve; CI, confidence interval.

causes of symptoms. As the original STONE score would generally suggest it is less likely for Black patients to have a kidney stone, it would actually lead toward a more thorough evaluation for kidney stones (and other causes of symptoms). That being said, we agree with the premise of the NEJM authors that in the absence of a compelling reason to include race, it should be revisited and revised when possible.

Since the NEJM piece, momentum has accelerated to remove race from clinical prediction rules when possible, especially when possibly harmful. This includes efforts by the American Academy of Pediatrics (AAP), the Coalition to End Racism in Clinical Algorithms (CERCA), the Doris Duke Foundation, and even the House Ways and Means Committee.[14]

In addition to race, since the original publication of the STONE score in 2014 there has been increased attention to the differences between sex assigned at birth and gender, and we are aware that the first element of both scores is binary and could be considered overly rigid. However, the data from the original stone score did not allow for separation of these data points, and addressing this aspect of the prediction rule was not an objective of this study.

It is problematic to include race as an element in predictive models without a more thorough and critical understanding of potential confounding variables. These may include underappreciated factors now recognized as the SDOH. These health-related social barriers can include varied lived experiences of economic stability, neighborhood and physical environment, educational opportunities, available food options, community, safety and social context, as well as access to and utilization of affordable high-quality and safe healthcare. We were unable to include these in a re-evaluation of the model as they were not completely collected as part of the original study. Future studies should strive to be vigilant in collecting and analyzing SDOH variables as crucial elements of patient presentation rather than simply using race as a blunt proxy.

The rationale for including race in our clinical prediction model, based on mathematical modeling from our data, was that it could improve accuracy and allow for tailored care across diverse populations. However, this approach failed to recognize that race lacks biological rigor, is societally determined, and is often in the eye of the beholder. It is also highly likely to be severely confounded with other uncollected and under-recognized socio-environmental variables. In the United States, patient categorization into Black race has been strongly associated with the legacy of racial discrimination in medical care.[15] For example, might Black patients be more likely to pass their stones at home and not seek ED care given access barriers or past negative experiences with the healthcare system?

We have shown that modifying a clinical prediction rule that formerly included race can be done without substantially impacting the tool's accuracy. The revised STONE score we propose performed similarly to the original STONE score. While further multicenter prospective validation is indicated, we recommend considering using the revised "STONE" score, without including race, for prediction of uncomplicated ureteral stone and risk stratification of alternate diagnoses.

AUTHOR CONTRIBUTIONS

Christopher L. Moore: Conceived and conducted original study; involved in all aspects of data reanalysis; drafting and revisions of manuscript; overall assumes responsibility for the manuscript. **Cary P. Gross**: Involved in original study; involved in design of

reanalysis; drafting; and revision of manuscript. **Louis Hart**: Involved in the design and reanalysis of data; and drafting of manuscript. **Annette M. Molinaro**: Performed the original and revised statistical analysis and participated in the manuscript development and revision. **Deborah Rhodes**: Involved in the reanalysis of the data; drafting and revision of manuscript. **Dinesh Singh**: Involved in the original study and participated in the drafting and revision of manuscript. **Cristiana Baloescu**: Participated in the reanalysis of data; drafting and revision of the manuscript.

ACKNOWLEDGEMENTS

The data used in this study were initially funded through a grant from the Agency for Healthcare Research and Quality (5R01HS018322-03) (however, there was no funding used for this reanalysis): Identifying unnecessary irradiation of patients with suspected renal colic. This funding provided resources to assist with the collection, management, analysis, and interpretation of the data. The AHRQ provided funding and oversight of funding, but was not directly involved in collection or cleaning of data, analysis of results, or drafting of the manuscript.

DATA AVAILABILITY STATEMENT

Deidentified/anonymized data used for both the retrospective and prospective portions of this study may be obtained by contacting Christopher Moore at chris.moore@yale.edu.

REFERENCES

1. Yearby R. Race based medicine, colorblind disease: how racism in medicine harms us all. *Am J Bioethics*. 2021;21:19-27.
2. Wright JL, Davis WS, Joseph MM, Ellison AM. Eliminating race-based medicine. *Pediatrics*. 2022;150(1):e2022057998.
3. Delgado C, Baweja M, Burrows NR, et al. Reassessing the inclusion of race in diagnosing kidney diseases: an interim report from the NKF-ASN task force. *J Am Soc Nephrol*. 2021;32:1305-1317. doi:10.1681/ASN.2021010039
4. Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight— reconsidering the use of race correction in clinical algorithms. *N Engl J Med*. 2020;383:874-882.
5. Khor S, Haupt EC, Hahn EE, Lyons LJL, Shankaran V, Bansal A. Racial and ethnic bias in risk prediction models for colorectal cancer recurrence when race and ethnicity are omitted as predictors. *JAMA Netw Open*. 2023;6:e2318495.
6. 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.
7. Michaels EK, Nakagawa Y, Miura N, Pursell S, Ito H. Racial variation in gender frequency of calcium urolithiasis. *J Urol*. 1994;152:2228-2231.
8. Sarmina I, Spirnak JP, Resnick MI. Urinary lithiasis in the black population: an epidemiological study and review of the literature. *J Urol*. 1987;138:14-17.
9. Cary MK. The racial incidence of urolithiasis. *J Urol*. 1937;37:651-654.
10. Rodgers AL. Race, ethnicity and urolithiasis: a critical review. *Urol Res*. 2013;41:99-103.
11. 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.e2.
12. Moore CL, Daniels B, Singh D, Luty S, Molinaro A. Prevalence and clinical importance of alternative causes of symptoms using a renal colic computed tomography protocol in patients with flank or back pain and absence of pyuria. *Acad Emerg Med*. 2013;20:470-478.
13. Sullivan LM, Massaro JM, D'Agostino RB. Presentation of multivariate data for clinical use: the Framingham Study risk score functions. *Stat Med*. 2004;23:1631-1660.
14. Kuehn BM. Citing harms, momentum grows to remove race from clinical algorithms. *JAMA*. 2024;331(6):463-465. doi:10.1001/jama.2023.25530
15. Roberts D. *Fatal Invention: How Science, Politics, and Big Business Re-Create Race in the Twenty-First Century*. The New Press; 2012.

How to cite this article: Moore CL, Gross CP, Hart LS, et al. Construction and performance of a clinical prediction rule for ureteral stone without the use of race or ethnicity: A new STONE score. *JACEP Open*. 2024;5:e13324. https://doi.org/10.1002/emp2.13324

AUTHOR BIOGRAPHY

Christopher L. Moore, MD, is a Professor in the Department of Emergency Medicine at Yale School of Medicine in New Haven, Connecticut. He is Chief of the Section of Emergency Ultrasound.