American Journal of Emergency Medicine 34 (2016) 230–234

Contents lists available at ScienceDirect

American Journal of Emergency Medicine

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

Original Contribution

Validity of STONE scores in younger patients presenting with suspected uncomplicated renal colic☆,☆☆,★

Elizabeth M. Schoenfeld, MD [a,*], Kye E. Poronsky, MA [a], Tala R. Elia, MD [a], Gavin R. Budhram, MD [a], Jane L. Garb, MS [b], Timothy J. Mader, MD [a]

[a] Department of Emergency Medicine, Baystate Medical Center/Tufts School of Medicine, Springfield, MA
[b] Epidemiology/Biostatistics, Department of Academic Affairs, Baystate Medical Center/Tufts School of Medicine, Springfield, MA

ARTICLE INFO

Article history:
Received 26 August 2015
Received in revised form 21 October 2015
Accepted 21 October 2015

ABSTRACT

Objectives: Recent studies have cast doubt on the routine need for emergent computed tomographic (CT) scan in patients with suspected renal colic. A clinical prediction rule, the STONE score, was recently published with the goal of helping clinicians predict obstructive kidney stones in noninfected flank pain patients before CT scan. We sought to examine the validity of this score in younger, noninfected flank pain patients.

Methods: A secondary analysis of a retrospective cohort study was performed to determine the validity of STONE scores for predicting the outcome of obstructive kidney stone in patients age 18 to 50 years presenting with flank pain suggestive of uncomplicated ureterolithiasis. Validity was measured by calculation of the area under the curve of the receiver operating characteristic curve. Sensitivity, specificity, negative predictive value, positive predictive value, and ± likelihood ratios were calculated for various cutoff values.

Results: Of 134 patients who met inclusion criteria, 56.7% were female, average age was 37 years, and 52% had an obstructing kidney stone by CT scan. The receiver operating characteristic curve for the STONE score had an area under the curve of 0.87 (95% confidence interval, 0.80-0.93) and indicated that a cutoff of greater than or equal to 8 would have a sensitivity of 78.6%, specificity of 84.4%, negative predictive value of 78.3%, positive predictive value of 84.6%, and + likelihood ratio of 4.9.

Conclusions: This analysis suggests that the STONE score is valid in younger populations. It can aid in determining pretest probability and help inform conversations about the likelihood of the diagnosis of renal colic before imaging, which may be useful for decision making.

1. Introduction

Every year, as many as 1 million Americans seek emergency department (ED) care for the pain caused by an obstructing kidney stone [1]. Recent estimates suggest that there are more than half a million computed tomographic (CT) scans performed in the United States each year for suspected kidney stones [2]. Although 80% of kidney stones pass spontaneously, the large majority of patients receive a CT scan, adding to the national radiation burden [3-5]. Older studies have advocated that the risk of alternative diagnoses warrants immediate ED CT scanning [6], but recent studies have challenged this [7,8], and a large comparative effectiveness trial demonstrated clinical equivalence of an "ultrasound first" pathway in regard to missed diagnoses [9]. A clinical prediction rule to help risk stratify patients in regard to their likelihood of a kidney stone as the cause of their symptoms could potentially decrease CT scans as patients with a high probability of a kidney stone have a lower probability of an alternative diagnosis [7]. A rule would be particularly important for younger patients who are less likely to have dangerous alternative diagnoses and more likely to see harm from a CT scan [5].

Recently, a clinical prediction rule was derived and validated using factors obtained from history and urinalysis to help predict the diagnosis of obstructing kidney stone [7]. The rule takes into account 5 factors for a total of 13 possible points (Sex, Timing, Origin [race], Nausea/vomiting, Erythrocytes [urine red blood cells]) (Table 1). The authors then assigned risk categories based on the predicted probability of stone for a given number of points as derived from their regression model, grouping the scores into low-, medium-, and high-probability groups. They validated the score in a prospective cohort of similar patients.

We sought to examine the external validity of the STONE score in a cohort of younger patients, age 18 to 50 years, presenting with flank pain. We evaluated whether these same categories of scores (0-5, 6-9,

☆ Presented at SAEM National Meeting, Dallas, May 2014.
☆☆ Grant funding: Use of RedCap Database was funded by UL1 TR000073.
★ Conflicts of interest: None of the authors has any conflicts of interests to disclose.
* Corresponding author at: Baystate Medical Center, ED Admin Offices, 759 Chestnut Street, S5433, Springfield, MA 01199.
E-mail address: ElizSchoen@gmail.com (E.M. Schoenfeld).

http://dx.doi.org/10.1016/j.ajem.2015.10.036
0735-6757/

Table 1
STONE score calculation table [7]

Variable	Points
Sex	
Male	2
Female	0
Timing (duration of pain prior to presentation)	
<6 h	3
6-24 h	1
>24 h	0
Origin	
Race: Nonblack	3
Race: Black	0
Nausea and vomiting	
Both	2
Nausea alone	1
Neither	0
Erythrocytes	
Hematuria present	3
Hematuria absent	0
Total	0-13

10-13) had similarly probabilities in our sample. We also sought to determine the area under the curve (AUC) of the receiver operating characteristic (ROC) curve and the likelihood ratios (LRs) given by the best cutoffs, as determined by the ROC.

2. Methods

2.1. Study design and setting

This study was approved by the local institutional review board. This was a secondary analysis of data from a retrospective cohort of patients in a 12-month period in 2011 to 2012. The original study was designed to examine the rate of clinically important alternative diagnoses in non-infected flank pain patients age 18 to 50 years [8]. The STONE score became available after the initiation of the original study; however, all criteria needed to calculate STONE scores were included in the initial study design, and no changes were made to data collection.

This ED has a combined pediatric and adult volume of more than 110,000 visits per year. The ED is staffed by attending physicians, residents, and physician assistants, and CT is available 24 hours a day, 7 days a week, with attending radiologist interpretation available until midnight and resident preliminary interpretations for 8 hours overnight.

Data were collected on the 5 STONE criteria, as well as final diagnosis by clinician and by CT report. For the AUC calculations, CT report was considered positive if a kidney stone was seen in the collecting system, ureters, or bladder, consistent with a symptomatic kidney stone.

2.2. Subjects

Electronic medical records were queried for coded triage complaints of "flank pain" or "possible kidney stone" from June 2011 until May of 2012. Charts were then manually examined for inclusion and exclusion criteria. Inclusion criteria included flank pain, age 18 to 50 years, and CT abdomen/pelvis performed. Exclusion criteria included left without being seen, no physician note, painless hematuria, concurrent diagnosis of urinary tract infection (UTI)/pyelonephritis, trauma causing chief complaint (major or minor), pregnancy known or discovered during visit, end-stage renal disease on hemodialysis/peritoneal dialysis or kidney transplant, recent surgical or urological intervention (60 days), already seen for this episode of pain (and index visit captured), or no "flank pain" in physician note.

2.3. Methods and measurements

Attempts were made to mitigate the biases inherent in chart reviews [10-12]. Standardized data abstraction forms were created a priori, and all reviewers were trained in the use of the standardized data abstraction form by the principal investigator. Interrater reliability was measured and is reported below and has been previously published [8]. Final diagnoses as decided by the treating physician and final CT scan read were recorded separately, as they did not always agree. Basic demographics were collected on all included patients. Data abstractors were not blinded to the original study's hypothesis.

Although patients who did not receive a CT scan were excluded, data were collected on these patients. The rate of renal colic in the non-CT group was determined by clinician's diagnosis, and the rate of renal

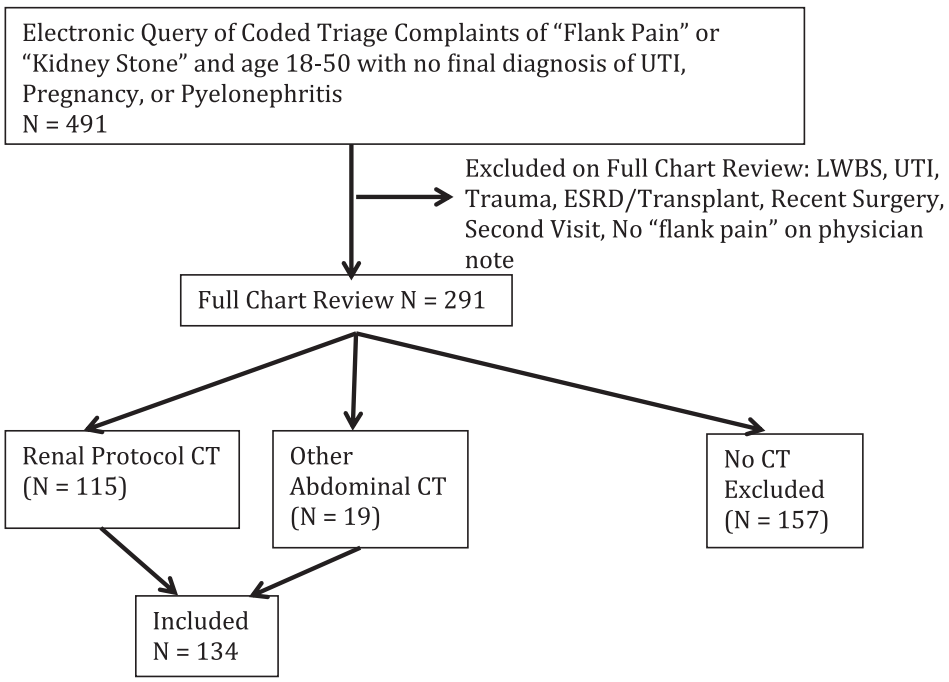

Fig. 1. Flow of inclusion of subjects.