

## RESEARCH

# Derivation and validation of a clinical prediction rule for uncomplicated ureteral stone—the STONE score: retrospective and prospective observational cohort studies

 OPEN ACCESS

Christopher L Moore *associate professor*<sup>1</sup>, Scott Bomann *emergency physician*<sup>1</sup>, Brock Daniels *emergency physician*<sup>2</sup>, Seth Luty *research assistant*<sup>3</sup>, Annette Molinaro *associate professor*<sup>1</sup>, Dinesh Singh *assistant professor*<sup>4</sup>, Cary P Gross *professor*<sup>5,6</sup>

<sup>1</sup>Department of Emergency Medicine, Yale University School of Medicine, New Haven, CT, USA; <sup>2</sup>Wellington Hospital, Wellington, New Zealand; <sup>3</sup>Yale New Haven Hospital, New Haven, CT, USA; <sup>4</sup>Departments of Neurosurgery and Epidemiology and Biostatistics, University of California San Francisco Medical School, San Francisco, CA, USA; <sup>5</sup>Department of Urology, Yale University School of Medicine, New Haven, CT, USA; <sup>6</sup>Department of Internal Medicine, Yale University School of Medicine, New Haven, CT, USA

## Abstract

**Objective** To derive and validate an objective clinical prediction rule for the presence of uncomplicated ureteral stones in patients eligible for computed tomography (CT). We hypothesized that patients with a high probability of ureteral stones would have a low probability of acutely important alternative findings.

**Design** Retrospective observational derivation cohort; prospective observational validation cohort.

**Setting** Urban tertiary care emergency department and suburban freestanding community emergency department.

**Participants** Adults undergoing non-contrast CT for suspected uncomplicated kidney stone. The derivation cohort comprised a random selection of patients undergoing CT between April 2005 and November 2010 (1040 patients); the validation cohort included consecutive prospectively enrolled patients from May 2011 to January 2013 (491 patients).

**Main outcome measures** In the derivation phase a priori factors potentially related to symptomatic ureteral stone were derived from the medical record blinded to the dictated CT report, which was separately categorized by diagnosis. Multivariate logistic regression was used to determine the top five factors associated with ureteral stone and these were assigned integer points to create a scoring system that was stratified into low, moderate, and high probability of ureteral stone. In the prospective phase this score was observationally derived blinded to

CT results and compared with the prevalence of ureteral stone and important alternative causes of symptoms.

**Results** The derivation sample included 1040 records, with five factors found to be most predictive of ureteral stone: male sex, short duration of pain, non-black race, presence of nausea or vomiting, and microscopic hematuria, yielding a score of 0-13 (the STONE score). Prospective validation was performed on 491 participants. In the derivation and validation cohorts ureteral stone was present in, respectively, 8.3% and 9.2% of the low probability (score 0-5) group, 51.6% and 51.3% of the moderate probability (score 6-9) group, and 89.6% and 88.6% of the high probability (score 10-13) group. In the high score group, acutely important alternative findings were present in 0.3% of the derivation cohort and 1.6% of the validation cohort.

**Conclusions** The STONE score reliably predicts the presence of uncomplicated ureteral stone and lower likelihood of acutely important alternative findings. Incorporation in future investigations may help to limit exposure to radiation and over-utilization of imaging.

**Trial registration** www.clinicaltrials.gov NCT01352676.

## Introduction

Kidney stones are estimated to occur at some point in nearly 1 in 11 people in the United States, with flank or kidney pain resulting in over two million annual visits to the emergency department.<sup>1,2</sup> Computed tomography (CT) has been described

Correspondence to: C L Moore [chris.moore@yale.edu](mailto:chris.moore@yale.edu)

Extra material supplied by the author (see <http://www.bmj.com/content/348/bmj.g2191?tab=related#datasupp>)

Patient factors assessed in deriving factors for score (greyed out factors were gathered for observational purposes but not considered for use in score)

Diagram for inclusion of patients in derivation (retrospective) phase

Enrollment diagram for validation (prospective) phase

as the “best imaging study to confirm the diagnosis of a urinary stone” and is now the first line imaging study for suspected kidney stone in the United States.<sup>3-5</sup> Though accurate, CT is costly, involves the use of ionizing radiation, and does not seem to have impacted patient centered outcomes, such as rates of diagnosis or hospital admission, in those with suspected kidney stones.<sup>3-7</sup>

Many patients with flank pain will not benefit from a CT scan, as most kidney stones will pass spontaneously. Moreover, it is unlikely that a CT scan in the setting of flank pain will detect acutely important alternative findings in patients without signs of infection.<sup>8</sup> Hence an objective clinical prediction rule for renal colic that could reliably identify patients highly likely to have a ureteral stone (and thus unlikely to have an important alternative diagnosis) may allow patients to be safely managed without imaging, or imaged with other approaches such as ultrasonography or reduced dose CT.

We derived and validated a clinical prediction score for ureteral stones that cause symptoms, identifying patients with either a very high or a very low probability of having an uncomplicated ureteral stone. We hypothesized that patients who are highly likely to have a kidney stone are unlikely to harbor an important alternative diagnosis, and may be appropriate for imaging choices other than standard dose CT.

## Methods

### Study design and setting

We performed a retrospective derivation and prospective validation of a clinical scoring system for ureteral stones that cause symptoms in two separate emergency departments with the same medical record systems.<sup>9-11</sup> The Yale New Haven Hospital emergency department is an urban, tertiary care teaching hospital and trauma center that sees over 80 000 adults annually. The Shoreline Medical Center emergency department is a freestanding eight bed suburban facility without residents, which sees approximately 20 000 adults and children annually. At the time of this study both sites utilized a templated, handwritten, scanned emergency department patient care record (Lynx Medical Systems, Bellevue, WA), with laboratory and dictated radiology reports on Sunrise Clinical Manage (Eclipsys, Atlanta, GA). The human investigation committee of the Yale institutional review board approved the derivation (retrospective) phase with a waiver of informed consent, and the validation (prospective) phase involved written informed consent from all patients.

### Derivation phase

We electronically retrieved the dictated reports of all patients receiving a CT “flank pain protocol” (the name given at both sites to a non-contrast enhanced CT protocol for suspected kidney stone) at either of the two emergency department sites between April 2005 and November 2010. Patients were eligible if the CT was performed in the emergency department and they were 18 years of age or older at the time of imaging. From an original set of over 5000 computed tomograms, we selected approximately one third of records (estimated to yield about 1000 records that met the inclusion criteria) for full record review using a random number spreadsheet function (Microsoft Excel, Redmond, WA). Exclusion criteria were lack of any flank or back pain, history of trauma, evidence of infection (subjective or objective fever or presence of leukocytes on urine dipstick analysis), known active malignancy, known renal disease (including creatinine >1.5 mg/dL or 133  $\mu$ mol/L), or previous urologic procedure (including lithotripsy or ureteral stent).<sup>8</sup>

### Power calculation—derivation and validation sets

Our selection of about 1000 records was based on pilot data and earlier studies indicating that about 50% of patients undergoing CT would have a ureteral stone, and about 20% of these would undergo intervention for ureteral stone (or 10% of overall population, about 100 patients). As a general rule, when using logistic regression, each independent element of a clinical prediction requires approximately 10 events.<sup>12</sup> This would have allowed us to incorporate a maximum of 10 elements in a rule to predict the need for intervention as well as being sufficiently powered to derive a rule for the more common outcome (any ureteral stone).

For the validation set we set minimally acceptable values for the classification probabilities of false and true positive fractions, of 0.05 and 0.95, respectively. All conclusions were to be based on a 90% ( $\alpha=0.1$ ) rectangular confidence region, using one sided exact confidence limits. As such we would attain 85% power with a minimum of 80 ureteral stones and a minimum of 256 non-stones.

### Data abstraction

Based on clinical experience and review of the literature, five physician co-investigators from three specialties (emergency medicine, internal medicine, and urology) identified an a priori list of factors thought to potentially be predictive of ureteral stone (see supplementary appendix 1). We conducted a literature review using key word searches in PubMed and relevant citations through Web of Science (Thomson Reuters). These factors were then abstracted from medical records blinded to CT reports.<sup>10</sup> The Lynx medical record used by emergency clinicians during the study period is a templated, handwritten chart that specifically prompts clinicians for the presence or absence of factors related to the chief complaint selected (typically flank or back pain), and was well suited to determining the presence or absence of factors. We abstracted the presence or absence of factors into a standardized form on an electronic database (Filemaker Pro 12, FileMaker, Santa Clara, CA).

We blindly abstracted and categorized the results of the dictated CT reports as previously described.<sup>8</sup> The reports were reviewed primarily to determine whether a kidney stone was causing symptoms or whether the computed tomogram showed another cause of symptoms. We considered a kidney stone to be the cause of symptoms if it was located from the renal pelvis to the ureterovesical junction (parenchymal stones were not considered to cause symptoms) or if signs of passed ureteral stone were specifically mentioned in the CT report. We also documented acutely important alternative causes of symptoms (such as appendicitis, diverticulitis, and others).<sup>8</sup> Other factors associated with kidney stone were also noted, including stone size, location, presence and degree of hydronephrosis or hydroureter, presence of perinephric or ureteral stranding, and asymptomatic stones as well as incidental findings (defined as unrelated to patient symptoms). We abstracted the CT results into a standard form on a separate FileMaker database.

### Inter-rater reliability

To determine inter-rater reliability of elements abstracted from the medical record, we blindly re-reviewed a subset of 50 randomly selected records. A priori, any element with a  $\kappa$  of below 0.6 was not eligible for inclusion in the prediction rule. We performed inter-rater reliability of categorization of CT scan results from a random selection of records.<sup>8</sup>

## Constructing the scoring system

All variables included were considered through univariate logistic regression analysis, with estimation of prevalence and odds ratios with corresponding 95% confidence intervals. We performed multivariate logistic regression, employing forward selection and 10-fold cross validation for model selection including estimation of two measures of prediction accuracy: the misclassification rate and the area under the receiver operating characteristic curve (AUC). Misclassification is a measure of prediction error, and ranges from 0 to 1, with lower scores indicating fewer errors in prediction. AUC ranges from 0.5 to 1, with higher scores indicating better prediction. The best model was the one that had a low cross validated misclassification rate and a high AUC. Subsequently, we included all observations to provide the most accurate estimates of the coefficients for the selected model and to derive a corresponding integer scoring system following the methods used in the Framingham study.<sup>11</sup> The simplicity of this scoring system allows a patient's risk to be calculated without the need for a calculator. Initially, we organized variables in the final multivariate model into meaningful categories, each with a specific reference value. We then assigned a referent risk for each factor with the base risk assigned 0 points in the scoring system, such that a higher point total conveys more risk. Next, we calculated the difference in terms of regression units between each category and the corresponding base category. We set the constant, B, as the number of regression units that corresponds to 1 point. We then computed the points for each risk factor's risk categories as the difference in regression units between each category and its base category divided by B. Subsequently we calculated the risk associated with each point total through the multiple logistic regression equation. We used a weighted  $\kappa$  test is used to verify the agreement between risk estimates based on the point system and those based on the multivariate logistic regression model. In addition to estimating AUC for summarizing the model's discrimination, we used the Hosmer and Lemeshow test to test for goodness of fit and calibration.

While the odds ratios (coefficients) from the multivariate regression analysis can be used to estimate the probability of an event (in this case ureteral stone), we sought to construct a more straightforward scoring system for clinical use without the use of complicated calculations. We assigned integer points to the presence of risk factors for ureteral stone using the coefficients from a multivariate analysis based on all observations, as described in the methods used to estimate the risk of cardiovascular disease in the Framingham study.<sup>11</sup> We computed points for each factor as the difference in regression units between each category and its base category, which was given a value of zero.

To assess the difference in accuracy between the integer point system and the logistic regression model we calculated the misclassification rate, AUC, and weighted  $\kappa$  based on differences in classification for each model. In addition to estimating AUC for summarizing the model's discrimination, we used the Hosmer and Lemeshow test to determine the 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) based on estimated clinical utility for the probability of ureteral stone by point total in each category.

## Prospective validation

From May of 2011 to February of 2013, consecutive patients presenting during defined periods to the emergency department sites in whom the clinician intended to obtain a CT scan for kidney stone were approached for enrollment. Both clinicians and enrolling staff were not aware of the specific elements of the rule derived in the retrospective phase. Defined enrollment shifts included overnights, weekends, and holidays, and an automatic paging system was set up to notify the research associate of all CTs ordered for renal colic. Review of the hospital imaging system was conducted daily to monitor any patients missed during enrollment or when enrollment was not taking place.

Before analysis of the validation data, the scoring system was developed from the derivation set as described previously, yielding a 0-13 point scale. Also before analysis of the prospective data we stratified this scale based on estimated clinical utility into low (about 10%), moderate (about 50%), and high (about 90%) probability of ureteral stone. Estimated clinical utility of cut points on the scale were arrived at through consensus of all investigators, including physicians from emergency medicine, internal medicine, and urology. Stratification into three groups enabled the derivation and validation sets to be compared for clinical utility for discrimination of risk as well as allowing estimates of the prevalence of more rare important alternative findings in each group.

The research associated recorded all relevant factors (listed in supplementary appendix 1) from the derivation phase for the enrolled patients before the results of the CT were known. Research associates were not aware of the elements of the STONE score when prospective data were collected. They assigned point values of 0-13 and category of risk to each patient in the validation cohort blinded to the CT result, and the CT result was categorized blinded to the clinical factors (except laterality of pain) and point total. We used bootstrapping to estimate Hosmer-Lemeshow test and discrimination (AUC) with AUC point estimates and 95% confidence intervals.

## Results

### Derivation sample

Of 5383 "flank pain protocol" CT scans (that is, the name for a non-enhanced CT scan using a renal colic protocol at our institution) performed in the emergency departments on patients 18 years of age or older during the retrospective period, 1853 (34.4%) were randomly selected for full record review. Of these, 1040 were complete records with no exclusion criteria (figure 1, also see supplementary figure 1). Table 1 lists the characteristics of the derivation and validation cohorts.

Approximately half (49.5%; 515 of 1040) of the patients had a ureteral stone that was causing symptoms on their computed tomogram, whereas 2.9% (30 of 1040) had acutely important alternative causes of symptoms. Inter-rater reliability for categorization of the CT result yielded a  $\kappa$  of 0.75-0.80, indicating excellent agreement. Table 2 shows the factors that were significant for the presence or absence of ureteral stone on univariate analysis.

### STONE score

Multivariate analysis yielded five factors that were most significantly associated with the presence of a ureteral stone: male sex, acute onset of pain, non-black race, presence of nausea or vomiting, and microscopic hematuria (table 3). Previous

visits to an emergency department were also significantly associated with a lower probability of ureteral stone but, to maximize generalizability between centers, were not included in the model. These five factors were incorporated into the STONE score with associated integer point values (table 3), yielding a total score ranging from 0-13.<sup>11</sup> The multivariate logistic regression model had a misclassification rate of 0.23 (95% confidence interval 0.22 to 0.23) and an AUC of 0.86 (95% confidence interval 0.79 to 0.93), whereas the STONE score had a misclassification rate of 0.23 (0.22 to 0.23) and an AUC of 0.82 (0.74 to 0.90). Agreement between the risk estimates based on the STONE score and those based on the multivariate logistic regression model demonstrated a weighted  $\kappa$  of 0.87 (95% confidence interval 0.86 to 0.87), indicating minimal loss of accuracy by assigning integer points to the factors.

### Prospective validation

From 25 May 2012 to 24 January 2013, 491 patients without exclusion criteria were enrolled (see supplementary figure 2). The characteristics of patients approached did not differ significantly from those that were not approached (table 1). For the validation cohort, the STONE score grouped into three levels of risk had an AUC of 0.792 (95% confidence interval 0.756 to 0.828) and the Hosmer-Lemeshow  $\chi^2=1.95$  was not significant ( $P=0.38$ ), indicating good discrimination and calibration.

### Comparison of derivation and validation sets

In the derivation and validation sets, respectively, 19.8% and 15.5% of patients were classified as having a low probability of kidney stone, 49.6% and 46.8% as moderate, and 30.6% and 37.7% as high. The prevalence of ureteral stone by group in the derivation and validation sets was, respectively, 8.3% and 9.2% in the low probability group, 51.6% and 51.3% in the moderate group, and 89.6% and 88.6% in the high group (figure). Overall, acutely important alternative causes of symptoms were found on CT scan in 2.9% and 3.7% of the derivation and validation cohorts, with acutely important alternative causes in 0.3% and 1.6% of the high probability group, respectively. Table 4 shows the causes and frequency of acutely important alternative findings in the overall derivation and validation sets.

### Discussion

This study showed that a clinical scoring system accurately predicts the likelihood of ureteral stone, which is inversely associated with likelihood of an acutely important alternate cause of symptoms. To our knowledge this is the first clinical scoring system to be derived and validated for prediction of uncomplicated ureteral stone in patients attending emergency departments in whom CT imaging is deemed indicated. A previous study from the intravenous pyelography era derived factors from 203 patients and validated the findings in 73 patients, finding four elements to be predictive of ureteral stone: flank pain, hematuria, acute onset of pain, and positive findings on a plain radiograph.<sup>13</sup> Our data show that the quantitative effects of the five factors incorporated into the STONE score can accurately predict ureteral stone and allow stratification of patients in the emergency department with suspected kidney stone into one of three groups: low probability ( $\leq 10\%$  chance of stone), moderate probability (about 50% chance of stone), and high probability (about 90% chance of stone).

Additionally, we found that the likelihood of an acutely important alternative finding is inversely proportional to the probability of a ureteral stone being present, as predicted by the

STONE score. While the overall presence of acutely important alternative findings was 2.9% in the derivation set and 3.8% in the validation set, the prevalence of clinically important alternative diagnoses in the high probability group was less than half of this: 0.3% and 1.6% in the derivation and validation cohorts, respectively.

### Clinical and policy implications

In deriving and validating this clinical prediction rule (rather than a decision rule), we are not necessarily stating that patients with a high stone score should not undergo CT imaging—though this may not be an unreasonable approach in certain situations. In any clinical situation the risk of a test (in this case from exposure to radiation) and the resources required to do the test will need to be balanced against the tolerance for uncertainty and risk of misdiagnosis on the part of both the clinician and the patient. In some patients—perhaps particularly younger ones who are more susceptible to radiation and less likely to have certain diagnoses such as diverticulitis, aortic disease, or malignancy—this score may be used to provide objective data to help balance the cost and risk of performing a CT. The other possibility is that this clinical prediction rule could be used to determine which patients may be most appropriate for substantially reduced dose CT, which has been shown to reliably identify ureteral stones, particularly large ones that may require intervention.<sup>14</sup>

### CT use in the United States, and public health implications

Since the landmark paper by Smith and colleagues in 1996, CT has become the first line test for kidney stone in the United States.<sup>3-15</sup> However, despite a 10-fold increase in the utilization of CT scanning for diagnosis of kidney stone from 1996-2007, the proportion of patients with a diagnosis of kidney stone, findings of significant alternative diagnoses, or hospital admission has not changed.<sup>3-7</sup> This suggests that the increase in CT use for diagnosis of this condition may not be substantially improving patient centered outcomes.<sup>16</sup> Outside of the United States, CT is not necessarily the first line test for suspected kidney stone.<sup>17-20</sup> In 2011 the European Urology Association released comprehensive guidelines on urolithiasis in which it stated that “ultrasonography should be used as the primary procedure.”<sup>21</sup> In 2007, the yearly rate of CT scanning in the United States was nearly 228 per 1000 population—more than double the rate in Canada and nearly four times the rate in the United Kingdom.<sup>22-23</sup> These data are not specific to imaging in kidney stones and do not include patient outcomes, but the presence of wide regional variation (particularly in a condition that is not life threatening) suggests an opportunity for more appropriate utilization.<sup>24-25</sup>

While the health risk attributable to a single CT scan is small, in a country of 310 million people (approximate US population) it is important to note a lifetime incidence of nephrolithiasis of approximately 10%.<sup>1</sup> If half of these people undergo a CT scan to detect nephrolithiasis (likely a conservative estimate as kidney stones are often recurrent and many patients undergo multiple CT scans<sup>26</sup>), we could expect 15 million CT scans to be performed on current US residents. In addition to the cost of this imaging, it could be estimated that exposure to ionizing radiation from CT would cause between 10 000 and 30 000 additional malignancies (using risk estimates of between 1 in 500 and 1 in 1500 for renal colic CT scans).<sup>27</sup>

In this setting CT was performed nearly as often in women as in men in both phases of the study (48.1% of CT scans in women

in the derivation phase; 44.4% in the validation phase). However, the diagnostic yield (percentage of patients with ureteral stones on CT) for men was much higher: 68.8% in the derivation phase and 66.7% in the validation phase compared with women (28.7% and 41.7%, respectively). The lower diagnostic yield in women coupled with a higher risk from radiation of the pelvis with CT suggests that women (especially younger women) may be a group that could benefit from more judicious use of CT radiation.

### Use of the score to select appropriate patients for reduced dose CT or ultrasonography

In terms of potential clinical utility, if a CT scan is being considered for suspected kidney stone and a patient has a high STONE score (which occurred in about a third of patients: 30.6% in the derivation cohort and 37.7% in the validation cohort), then the patient is very likely to have a kidney stone and very unlikely to have an important non-kidney stone cause of symptoms. Thus, if the STONE score is high a CT might be avoided entirely or a reduced dose CT could be performed (to ensure that there is not a large stone that may require intervention). It is important to note that it is still possible to miss an important alternative diagnosis in the high probability group if CT is not performed (of the roughly 10% of patients in the high group, about 10% of these, or 1-2% of the overall group, had an important alternative finding). However, the STONE score offers objective data to both the clinician and the patient that could help guide shared decision making about CT scanning, which is not without risk in terms of radiation and incidental findings that may lead to further testing or intervention. Our hope is that this score can be incorporated into imaging decisions for suspected renal colic to decrease exposure to radiation and over-utilization of imaging (that is, imaging without improvement in patient care).<sup>28</sup> Further investigation, potentially including a randomized trial, may help to elucidate this.

Most kidney stones (smaller stones, about 80% in this study as is generally the case) will pass spontaneously with treatment of the symptoms. Patients with a very high probability of ureteral stone thus may not require any imaging and could be managed with pain control and drugs to enhance stone expulsion, with definitive diagnosis using a urine strainer. Clinicians may, however, still want to perform a CT to exclude potentially serious alternative causes of symptoms<sup>29</sup> and to determine the size and location of any stone (with implications for prognosis and intervention).<sup>30</sup> In this case, patients with a high STONE score may be ideally suited for substantially reduced dose CT scanning. Though data on low dose protocols have been published outside of the United States<sup>31-33</sup> and the American College of Radiology states reduced dose techniques are “preferred,”<sup>34</sup> data from the Dose Imaging Registry (part of the American College of Radiology National Radiology of Data Registry: [www.nrdcr.acr.org](http://www.nrdcr.acr.org)) indicates that the mean institutional dose for CT for renal colic is still greater than 10 mSv, and reduced dose techniques are rarely used in US hospitals (in press).<sup>34</sup>

Reduced dose CT has been shown to be accurate for kidney stones, particularly larger ones that may require intervention, but has not been widely used in the United States, likely because of concerns about accuracy in an unselected population.<sup>14 34</sup> Reluctance to implement reduced dose CT protocols for renal colic may result from fear of missing other disease. An investigator looking at reduced dose CT for renal colic noted that to put these reduced dose protocols into practice they “would want to target it at patients who have a high pretest

probability of calculi and obstructive uropathy, since the ability to detect other pathology is hindered.”<sup>35</sup> In addition to predicting kidney stone, our data show that the group that is most likely to have kidney stones is also unlikely (<2%) to have an important alternative cause of symptoms. A probability of disease under 2% has been identified as a testing threshold (point at which the negatives of a test outweigh the positives) for CT use in detecting other important diseases, such as pulmonary embolism.<sup>36</sup> Identifying patients in this group could safely direct some patients with suspected kidney stone to low dose or ultra low dose CT.

Ultrasound is another option that may be used for imaging in suspected renal colic, and ultrasonography is often a first line test outside of the United States.<sup>19 21</sup> It has the advantage of avoiding radiation entirely and is sometimes definitively diagnostic: identifying the presence, size, and location of a kidney stone that is causing symptoms. Often, however, ultrasonography may show indirect evidence of obstruction (hydronephrosis) without visualizing the actual ureteral stone, which may be obscured by bowel. We did find the presence of hydronephrosis on CT to be highly predictive of ureteral stone, and future work will incorporate the presence of hydronephrosis on ultrasonography into the STONE score.

At our institution, the STONE score has been incorporated into the computerized physician order entry system (Epic, Verona WI). When a clinician orders a CT for kidney stone the questions asked and a STONE score with risk category accompanies the radiology order. This has been welcomed by the radiologists who were often unsure of the perceived likelihood of kidney stone on the part of the ordering physician. We have found that the STONE score is easily entered and calculated using our electronic health record. We are also currently using the STONE score in a prospective study to select patients who are appropriate for either expectant management (no CT) or an ultra low dose CT, with a radiation dose that is about 90% lower than conventional CT (effective dose of around 1 mSv, about that of a plain abdominal radiograph). On a population basis, assuming the no threshold linear model suggested by the Biologic Effects of Ionizing Radiation report (currently BEIR VII), an equivalent reduction in cancer risk could be expected.<sup>37</sup> The current average effective dose of CT in the United States is 11.2 mSv, with only 2% of CT scans done using low doses.<sup>38</sup>

### Strengths and limitations of this study

An important limitation of this study is that gestalt clinician pretest probability for kidney stone (that is, the overall clinician estimate for likelihood of kidney stone after initial clinician evaluation) has not been thoroughly investigated, and it is possible that it would perform similarly to an objective clinical prediction rule. A study by Abramson and colleagues showed that the pretest probability of emergency department physicians obtaining CT for suspected kidney stone clustered in the 41-60% and 71-90% ranges.<sup>39</sup> However, the use of a relatively objective scoring system has the advantage that it is not dependent on clinician experience. In pulmonary embolism, for example, while gestalt pretest probability has been shown to be reasonably accurate, authors comparing gestalt pretest probability to objective scoring systems conclude that they “advocate the use of a clinical prediction rule because it has been shown to be accurate and can be used by less-experienced clinicians.”<sup>40</sup> This study is also limited by being derived and validated in the same clinical setting; it is not known how well it would perform in other settings.

## Conclusion

We have derived and validated a clinical prediction score for the presence of ureteral stones that cause symptoms. Multicenter validation and evaluation of incorporating the STONE score into imaging algorithms is warranted.

We thank Christal Esposito, project manager, and research associates Pasquale Cicarella and Richelle Jessey and other research associates for their help in collecting and organizing data. The authors have followed the STROBE checklist in collecting and reporting their data. Elements of the checklist are incorporated into the manuscript. Further details of any particular checklist item are available on request from the corresponding author (chris.moore@yale.edu).

Contributors: CLM conceived the work, monitored data collection for the whole trial, oversaw data cleaning and statistical analysis, and drafted and revised the manuscript. He is guarantor. SB helped with project conception and participated in data analysis and manuscript revision. BD assisted with data analysis and participated in manuscript revisions. SL was lead research assistant and data manager, he participated in data collection, designed data collection tools, oversaw cleaning and management of data, and assisted with presentation of results in the manuscript. AM provided input into methodology, performed statistical analyses, and participated in manuscript revisions. DS participated in study design, analysis of data, and manuscript revisions. CPG participated in study design, analysis of data, and manuscript revisions. The authors confirm that data were collected, results analyzed, and the manuscript prepared without influence from funding agencies.

Funding: This study was funded through a grant from the Agency for Healthcare Research and Quality (5R01HS018322-03): 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, and preparation and review of the manuscript. 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.

Competing interests: All authors have completed the ICMJE uniform disclosure form at [www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) and declare: All authors with the exception of BD and SB were funded in part by a research grant from the Agency for Healthcare Research and Quality (5R01HS018322-03); CPG received support from the Robert Wood Johnson clinical scholars program; CLM has received compensation as a consultant from Philips Healthcare and Sonosite (a subsidiary of FujiFilm); CPG is a scientific advisory board member for Fair Health; CPG receives funding from Medtronic as a collaborator on the Yale University Open Data Access project; there are no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: This study was approved by the Yale Human Investigation Committee (institutional review board).

Data sharing: The raw data from this study is not currently available online; however, the authors are willing to share a deidentified dataset of primary data with interested parties. Available on request from the corresponding author (chris.moore@yale.edu).

Transparency: CLM, the lead author and guarantor, affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

- 1 Scales CD, Smith AC, Hanley JM, Saigal CS. Prevalence of kidney stones in the United States. *Eur Urol* 2012;62:160-5. doi:10.1016/j.eururo.2012.03.052.
- 2 Pearle MS, Calhoun EA, Curhan GC. Urologic diseases in America project: urolithiasis. *J Urol* 2005;173:848-57. doi:10.1097/01.ju.0000152082.14384.d7.
- 3 Hyams ES, Korley FK, Pham JC, Matlaga BR. Trends in imaging use during the emergency department evaluation of flank pain. *J Urol* 2011;186:2270-4. doi:10.1016/j.juro.2011.07.079.

- 4 Imaging EP on U. American College of Radiology appropriateness criteria: acute onset flank pain—suspicion of stone disease. 2011. [www.acr.org/-/media/ACR/Documents/AppCriteria/Diagnostic/AcuteOnsetFlankPainSuspicionStoneDisease.pdf](http://www.acr.org/-/media/ACR/Documents/AppCriteria/Diagnostic/AcuteOnsetFlankPainSuspicionStoneDisease.pdf).
- 5 Teichman JMH. Acute renal colic from ureteral calculus. *N Engl J Med* 2004;350:684-93.
- 6 Gottlieb RH, Thao CL, Erturk EN, Sotack JL, Voci SL, Holloway RG, et al. CT in detecting urinary tract calculi: influence on patient imaging and clinical outcomes. *Radiology* 2002;225:441-9.
- 7 Westphalen AC, Hsia RY, Maselli JH, Wang R, Gonzales R. Radiological imaging of patients with suspected urinary tract stones: national trends, diagnoses, and predictors. *Acad Emerg Med* 2011;18:699-707. doi:10.1111/j.1553-2712.2011.01103.x.
- 8 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;20:470-8. doi:10.1111/acem.12127.
- 9 Stiell IG, Wells GA. Methodologic standards for the development of clinical decision rules in emergency medicine. *Ann Emerg Med* 1999;33:437-47. [www.ncbi.nlm.nih.gov/pubmed/10092723](http://www.ncbi.nlm.nih.gov/pubmed/10092723).
- 10 Gilbert EH, Lowenstein SR, Koziol-McLain J, Barta DC, Steiner J. Chart reviews in emergency medicine research: where are the methods? *Ann Emerg Med* 1996;27:305-8. [www.ncbi.nlm.nih.gov/pubmed/8599488](http://www.ncbi.nlm.nih.gov/pubmed/8599488).
- 11 Sullivan LM, Massaro JM, D'Agostino RB. Presentation of multivariate data for clinical use: the Framingham Study risk score functions. *Stat Med* 2004;23:1631-60. doi:10.1002/sim.1742.
- 12 Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules—applications and methodological standards. *N Engl J Med* 1985;313:793-9.
- 13 Elton TJ, Roth CS, Berquist TH, Silverstein MD. A clinical prediction rule for the diagnosis of ureteral calculi in emergency departments. *J Gen Intern Med* 1993;8:57-62. [www.ncbi.nlm.nih.gov/pubmed/8441076](http://www.ncbi.nlm.nih.gov/pubmed/8441076).
- 14 Niemann T, Kollmann T, Bongartz G. Diagnostic performance of low-dose CT for the detection of urolithiasis: a meta-analysis. *Am J Roentgenol* 2008;191:396-401. doi:10.2214/AJR.07.3414.
- 15 Smith RC, Verga M, McCarthy S, Rosenfield AT. Diagnosis of acute flank pain: value of unenhanced helical CT. *Am J Roentgenol* 1996;166:97-101.
- 16 Swensen SJ. Patient-centered Imaging. *Am J Med* 2012;125:115-7. doi:10.1016/j.amjmed.2011.06.002.
- 17 Paltas M, Farkas A, Fisher D, Zagal I, Hadas-Halpern I. Ultrasound vs CT for the detection of ureteric stones in patients with renal colic. *Br J Radiol* 2001;74:901-4. [www.ncbi.nlm.nih.gov/pubmed/11675305](http://www.ncbi.nlm.nih.gov/pubmed/11675305).
- 18 Park SJ, Yi BH, Lee HK, Kim YH, Kim HC. Evaluation of patients with suspected ureteral calculi using sonography as an initial diagnostic tool. *J Ultrasound Med* 2008;27:1441-50.
- 19 Moore CL, Scoutt L. Sonography first for acute flank pain? *J Ultrasound Med* 2012;31:1703-11. [www.ncbi.nlm.nih.gov/pubmed/23091240](http://www.ncbi.nlm.nih.gov/pubmed/23091240).
- 20 Ripollés T, Agramunt M, Errando J, Martínez MJ, Coronel B, Morales M. Suspected ureteral colic: plain film and sonography vs unenhanced helical CT. A prospective study in 66 patients. *Eur Radiol* 2004;14:129-36. doi:10.1007/s00300-003-1924-6.
- 21 Tiselius HG, Ackermann D, Alken P, Buck C, Conort P, Gallucci M. Guidelines on urolithiasis. *Eur Urol* 2011;40:362-71. [www.ncbi.nlm.nih.gov/pubmed/19585096](http://www.ncbi.nlm.nih.gov/pubmed/19585096).
- 22 Rampell C. Using, and overusing, medical technologies. *New York Times*. 16 Dec 2009.
- 23 Berdahl CT, Vermeulen MJ, Larson DB, Schul MJ. Emergency department computed tomography utilization in the United States and Canada. *Ann Emerg Med* 2013;62:486-94.e3. doi:10.1016/j.annemergmed.2013.02.018.
- 24 Song Y, Skinner J, Bynum J, Sutherland J, Wennberg JE, Fisher ES. Regional variations in diagnostic practices. *N Engl J Med*. 2010;363:45-53. doi:10.1056/NEJMs0910881.
- 25 Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas F, Pinder EL. The implications of regional variations in medicare spending. Part 2: health outcomes and satisfaction with care. *Ann Intern Med* 2013;138:288-98.
- 26 Katz SI, Saluja S, Brink J, Forman HP. Radiation dose associated with unenhanced CT for suspected renal colic: impact of repetitive studies. *Am J Roentgenol* 2006;186:1120-4. doi:10.2214/AJR.04.1838.
- 27 Smith-Bindman R, Lipson J, Marcus R, Kim KP, Mahesh M, Gould R, et al. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. *Arch Intern Med* 2009;169:2078-86. doi:10.1001/archinternmed.2009.427.
- 28 Brenner DJ. Minimising medically unwarranted computed tomography scans. *Ann ICRP* 2012;41:161-9. doi:10.1016/j.icrp.2012.06.004.
- 29 Hoppe H, Studer R, Kessler TM, Vock P, Studer UE, Thoeny HC. Alternate or additional findings to stone disease on unenhanced computerized tomography for acute flank pain can impact management. *J Urol* 2006;175:1725-30; discussion 1730. doi:10.1016/S0022-5347(05)00987-0.
- 30 Papa L, Stiell IG, Wells G, Ball I, Battram E, Mahoney JE. Predicting intervention in renal colic patients after emergency department evaluation. *CJEM* 2005;7:78-86. [www.ncbi.nlm.nih.gov/pubmed/17355656](http://www.ncbi.nlm.nih.gov/pubmed/17355656).
- 31 Poletti P-A, Platon A, Rutschmann OT, Schmidlin FR, Iselin CE, Becker CD. Low-dose versus standard-dose CT protocol in patients with clinically suspected renal colic. *Am J Roentgenol* 2007;188:927-33. doi:10.2214/AJR.06.0793.
- 32 Kluner C, Hein PA, Gralla O, Hein E, Hamm B, Romano V, et al. Does ultra-low-dose CT with a radiation dose equivalent to that of KUB suffice to detect renal and ureteral calculi? *J Comput Assist Tomogr* 2006;30:44-50. [www.ncbi.nlm.nih.gov/pubmed/16365571](http://www.ncbi.nlm.nih.gov/pubmed/16365571).
- 33 Kim BS, Hwang IK, Choi YW, Namkung S, Kim HC, Hwang WC, et al. Low-dose and standard-dose unenhanced helical computed tomography for the assessment of acute renal colic: prospective comparative study. *Acta radiol* 2005;46:756-63. doi:10.1080/02841850500216004.
- 34 Lukasiewicz A, Weinreb J, Coombs LP, Bhargava M, Ghita M, Moore CL. Radiation dose index of CTs for kidney stone performed in the United States. *Acad Emerg Med* 2013;20:S54.
- 35 Frei R. Ultra-low CT dose accurately detects renal stones. *Ren Urol News* 2012. 2013. [www.renalandurologynews.com/ultra-low-ct-dose-accurately-detects-renal-stones/article/248315/](http://www.renalandurologynews.com/ultra-low-ct-dose-accurately-detects-renal-stones/article/248315/).
- 36 Kline J. Further illumination of the test threshold approach in the care of emergency department patients with symptoms of pulmonary embolism. *Ann Emerg Med* 2010;55:327-30. doi:10.1016/j.annemergmed.2010.01.005.
- 37 National Academy of Sciences. Health risks from exposure to low levels of ionizing radiation: BEIR VII—phase 2 (executive summary). 2006. [www.nap.edu/openbook.php?record\\_id=11340&page=1](http://www.nap.edu/openbook.php?record_id=11340&page=1).

**What is already known on this topic**

Kidney stones are common, and imaging with computed tomography (CT) is now the first line diagnostic test

However, CT has not been shown to improve patient centered outcomes

An objective, validated clinical prediction rule for uncomplicated ureteral stone has not been demonstrated and could help decrease exposure to radiation or over-utilization of imaging

**What this study adds**

A clinical prediction rule was derived and validated that can identify patients with a high probability of uncomplicated ureteral stone and absence of other important cause of symptoms

Results from this study may be used to select patients who could benefit from management without CT, or from reduced dose CT

- 38 Lukasciewicz A, Bhargavan-Chatfield M, Coombs L, Ghita M, Weinreb J, Gunabushanam G, et al. Radiation dose index of renal colic protocol CT studies in the United States: a report from the American College of Radiology National Radiology Data Registry. *Radiology* 2014. <http://pubs.rsna.org/doi/full/10.1148/radiol.14131601>.
- 39 Abramson S, Walders N, Applegate KE, Gilkeson RC, Robbin MR. Impact in the emergency department of unenhanced CT on diagnostic confidence and therapeutic efficacy in patients with suspected renal colic. *Am J Roentgenol* 2000;175:1689-95.
- 40 Chunilal SD, Eikelboom JW, Attia J, Miniati M, Panju AA, Simel DL, et al. Does this patient have pulmonary embolism? *JAMA* 2003;290:2849-58.

**Accepted:** 7 March 2014

Cite this as: *BMJ* 2014;348:g2191

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>.

## Tables

**Table 1 | Demographics of derivation and validation cohorts. Values are numbers (percentages) unless stated otherwise**

Characteristics	Derivation cohort (n=1040)	Validation cohort (n=491)
Mean (SD) age (years)	44.8 (14.9)	45.8 (14.7)
Female sex	501 (48.1)	218 (44.4)
Race:		
White	883 (84.9)	411 (83.7)
Black	110 (10.6)	57 (11.6)
Other	47 (4.5)	23 (4.7)
Location of enrolment:		
Yale-New Haven Hospital ED	722 (69.4)	357 (72.7)
Shoreline Medical Center ED	318 (30.6)	134 (27.3)
Cause of symptoms on CT:		
Symptomatic ureteral stone	515 (49.5)	274 (55.8)
Acutely important alternative cause	30 (2.9)	18 (3.7)
Disposition:		
Admit	71 (6.8)	52 (10.6)
Discharge	969 (93.2)	439 (89.4)

ED=emergency department; CT=computed tomography.



**Table 2 | Significant predictors for presence or absence of ureteral stone (univariate analysis of derivation set), with odds ratios and 95% confidence intervals**

Factors	Baseline (odds ratio 1.0)	No (%) of total with factor	No (%) of those with factor with ureteral stone	Odds ratio (95% CI)
<b>Personal characteristics:</b>				
Male sex	Female sex	539 (51.8)	371 (68.8)	3.4 (2.6 to 4.4)
Non-black race	Black race	930 (89.4)	547 (58.8)	6.1 (3.7 to 9.9)
Arrival by ambulance	Arrival by other mode	156 (15.0)	96 (61.5)	1.4 (1.0 to 2.0)
<b>History of present illness:</b>				
Any flank pain	No flank pain	973 (93.6)	551 (56.6)	3.8 (2.2 to 6.9)
Any back pain	No back pain	315 (30.3)	134 (42.5)	0.5 (0.4 to 0.6)
Symptoms lateralized	Symptoms non-lateralized	853 (82.0)	480 (56.3)	1.5 (1.1 to 2.0)
Pain onset abrupt or sudden	Pain onset gradual or unknown	630 (61.0)	420 (66.7)	3.5 (2.7 to 4.6)
Pain course constant	Pain course not constant	367 (35.3)	223 (60.8)	1.5 (1.1 to 1.9)
Pain with movement	No pain with movement	222 (21.3)	91 (41)	0.5 (0.4 to 0.7)
Pain duration <6 hours	Pain course 1 day to 1 week	375 (36.1)	292 (77.9)	5.8 (4.1 to 8.2)
Pain duration 6 h-1 day	Pain course 1 day to 1 week	259 (24.9)	137 (52.9)	1.8 (1.3 to 2.6)
Pain >1 week	Pain course 1 day to 1 week	113 (10.9)	23 (20.4)	0.4 (0.2 to 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 to 2.8)
Radiation of pain to groin	No radiation of pain to groin	336 (32.3)	229 (68.2)	2.3 (1.8 to 3.0)
Nausea alone	No nausea or vomiting	311 (29.9)	176 (56.6)	1.9 (1.4 to 2.6)
Nausea with vomiting	No nausea or vomiting	298 (28.7)	219 (73.5)	4.1 (3.0 to 5.7)
Presence of diarrhea	Absence of diarrhea	53 (5.1)	21 (39.6)	0.5 (0.3 to 0.9)
Presence of dysuria	Dysuria not present	211 (20.3)	129 (61.1)	1.9 (1.5 to 2.5)
Subjective hematuria	Subjective hematuria not present	205 (19.7)	139 (67.8)	2.0 (1.5 to 2.8)
<b>Medical, family, and social history:</b>				
Presence of any allergy	No allergy present	335 (32.2)	143 (42.7)	0.5 (0.4 to 0.6)
No prior visits to emergency department documented	Prior visits to emergency department documented	592 (56.9)	404 (68.2)	3.7 (2.9 to 4.8)
Family history of kidney stones	No family history or not mentioned	63 (6.1)	50 (79.4)	3.4 (1.9 to 6.6)
Any history of smoking	No history of smoking	195 (18.8)	78 (40)	0.5 (0.4 to 0.7)
History of kidney stones	No history of kidney stones	326 (31.3)	194 (59.5)	1.4 (1.0 to 1.7)
Any surgical history	No surgical history	302 (29)	141 (46.7)	0.6 (0.5 to 0.8)
Taking any drugs	No drugs documented	464 (44.7)	227 (48.9)	0.7 (0.5 to 0.8)
<b>Physical examination:</b>				
Raised systolic blood pressure, each 10 mm Hg	Mean 134 (SD 35) mm Hg	NA	NA	1.2 (1.1 to 1.2)
Raised diastolic blood pressure, each 10 mm Hg	Mean 85 (SD 13) mm Hg	NA	NA	1.3 (1.2 to 1.4)
Raised pulse, per 10 beats/min	Mean 83 (SD 15) beats/min	NA	NA	0.8 (0.8 to 0.9)
Right lower quadrant tenderness	No right lower quadrant tenderness	171 (16.4)	107 (62.6)	1.5 (1.1 to 2.1)
Right or left lower quadrant tenderness	No right or left lower quadrant tenderness	330 (31.7)	198 (60.0)	1.4 (1.1 to 1.8)
Upper abdominal tenderness	No upper tenderness	91 (8.8)	38 (41.8)	0.6 (0.4 to 0.9)
Lumbar or back tenderness		106 (10.2)	36 (33.0)	0.4 (0.2 to 0.6)
<b>Laboratory values:</b>				
Any erythrocytes in urine	No erythrocytes in urine	717 (68.9)	473 (66.0)	4.7 (3.5 to 6.2)
Creatinine, each 8.84 µmol/L (0.1 mg/dL) increase	88.4 (SD 35.4) µmol/L (1.0 SD 0.4 mg/dL)	NA	NA	0.013 (0.012 to 0.014)

NA=not applicable.

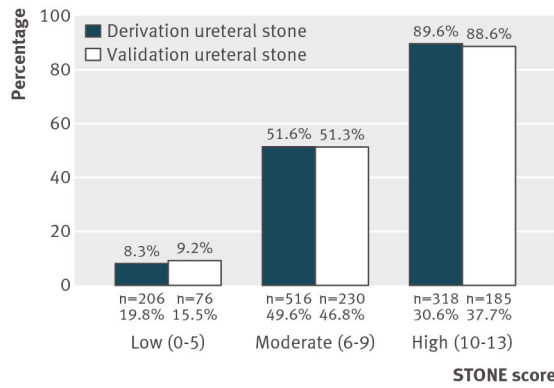
Table 3| STONE score, factors, and categories

STONE score by factors and categories	Odds ratio (95% CI)	Points
<b>Sex</b>		
Sex:		
Female	1	0
Male	4.31 (3.13 to 5.98)	2
<b>Timing</b>		
Duration of pain to presentation:		
>24 hours	1	0
6-24 hours	1.85 (1.27 to 2.70)	1
<6 hours	6.34 (4.26 to 9.33)	3
<b>Origin</b>		
Race:		
Black	1	0
Non-black	6.77 (3.79 to 12.64)	3
<b>Nausea</b>		
Nausea and vomiting:		
None	1	0
Nausea alone	1.98 (1.38 to 2.86)	1
Vomiting alone	5.26 (3.53 to 7.93)	2
<b>Erythrocytes</b>		
Hematuria (on urine dipstick):		
Absent	1	0
Present	5.61 (3.96 to 8.04)	3
Total		0-13

**Table 4 | Types and frequency of acutely important alternative causes of symptoms in derivation and validation sets, listed by decreasing frequency in derivation set**

Acutely important alternative cause of symptoms	Derivation set (n=1040)	Validation set (n=491)
Diverticulitis	6	4
Appendicitis	5	4
Malignancy or concerning mass	4	1
Ovarian or adnexal cause	4	1
Pyelonephritis	3	1
Ruptured angiomyolipoma	2	0
Cholecystitis	1	2
Pneumonia	1	1
Retroperitoneal fibrosis	1	0
Perforated viscus	1	0
Bowel obstruction	1	1
Colitis	1	1
Aortic aneurysm	0	1
Pancreatitis	0	1
Total No (%)	30 (2.9%)	18 (3.8%)

### Figure



Prevalence of ureteral stone by STONE score category in derivation and validation cohorts. Percentages at top of bars indicate prevalence of ureteral stone in group. Values under bars indicate number within derivation and validation sets that fell within risk stratum