HHS Public Access

Author manuscript

Am J Emerg Med. Author manuscript; available in PMC 2018 April 01.

Published in final edited form as:

Am J Emerg Med. 2017 April ; 35(4): 554–563. doi:10.1016/j.ajem.2016.12.009.

Derivation of decision rules to predict clinically important outcomes in acute flank pain patients

Ralph C. Wang[a,*], **Robert M. Rodriguez**[a], **Jahan Fahimi**[a], **M. Kennedy Hall, MD, MHS**[b], **Stephen Shiboski**[c], **Tom Chi**[d], and **Rebecca Smith-Bindman**[c,e]

[a]Department of Emergency Medicine, University of California, San Francisco, San Francisco, CA, USA

[b]Division of Emergency Medicine, University of Washington, Seattle, WA, USA

[c]Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA

[d]Department of Urology, University of California, San Francisco, San Francisco, CA, USA

[e]Department of Radiology and Biomedical Imaging, University of California, San Francisco, CA, USA

Abstract

Objective—Routine CT for patients with acute flank pain has not been shown to improve patient outcomes, and it may unnecessarily expose patients to radiation and increased costs. As preliminary steps toward the development of a guideline for selective CT, we sought to determine the prevalence of clinically important outcomes in patients with acute flank pain and derive preliminary decision rules.

Methods—We analyzed data from a randomized trial of CT vs. ultrasonography for patients with acute flank pain from 15 EDs between October 2011 and February 2013. Clinically important outcomes were defined as inpatient admission for ureteral stones and alternative diagnoses. Clinically important stones were defined as stones requiring urologic intervention. We sought to derive highly sensitive decision rules for both outcomes.

Results—Of 2759 participants, 236 (8.6%) had a clinically important outcome and 143 (5.2%) had a clinically important stone. A CDR including anemia (hemoglobin <13.2 g/dl), WBC count >11 000/μl, age > 42 years, and the absence of CVAT had a sensitivity of 97.9% (95% CI 94.8–99.2%) and specificity of 18.7% (95% 17.2–20.2%) for clinically important outcome. A CDR

[*]Corresponding author at: Department of Emergency Medicine, 505 Parnassus Avenue, San Francisco, CA 94143, USA. ralph.wang@ucsf.edu (R.C. Wang).

Conflict of interest

The authors declare no conflicts of interest.

Author contributions

RCW conceived the work, performed data collection, statistical analysis, drafted and critically revised the manuscript. RR helped with study conception, design, data analysis, and participated in manuscript revision. MKH helped to perform the data analysis, manuscript preparation, and provided critical revisions. JF helped to perform data analysis, and critically revised the manuscript. SS provided methodological guidance and critically revised the manuscript. TC helped with study design and critically revised the manuscript. RSB helped with study conception, design, and participated in manuscript revision. All authors had full access to the data, take responsibility for the integrity of the data and have approved the manuscript. The data were collected, results analyzed, and the manuscript was prepared without influence from funding agencies.

including hydronephrosis, prior history of stone, and WBC count <8300/μl had a sensitivity of 98.6% (95% CI 94.5–99.7%) and specificity of 26.0% (95% 24.2–27.7%) for clinically important stone.

Conclusions—We determined the prevalence of clinically important outcomes in patients with acute flank pain, and derived preliminary high sensitivity CDRs that predict them. Validation of CDRs with similar test characteristics would require prospective enrollment of 2100 patients.

1. Introduction

1.1. Background

An estimated two million patients present to the emergency department (ED) for acute flank pain annually [1–3]. Currently, computed tomography (CT) scan is the most commonly used imaging test, valued for its excellent sensitivity and specificity for ureteral stone and its ability to detect important alternative diagnoses, such as appendicitis, diverticulitis, and abdominal aortic aneurysm [2,4]. However, CT scan for acute flank pain may be over-utilized: CT scan is obtained in 70% of ED visits for urolithiasis, but only 10% of patients presenting with acute flank pain are admitted for management of a clinically important outcome, defined ureteral stone requiring urologic intervention or an alternative (non-kidney stone) diagnosis requiring inpatient admission [2,5–7]. The dramatic rise in CT use for acute flank pain has not been shown to increase the rate of diagnosis of urolithiasis, alternative diagnosis, or hospitalization [3,4]. Also, indiscriminate CT use may lead to costly, inefficient care with significant associated harms. Experts have estimated that CT scan radiation may cause 3–5% of all future malignancies, and with radiation-vulnerable organs directly in the field, CT scan of the flank and abdomen may be especially risky. The dramatically increased CT scan use has fueled skyrocketing costs of care – fees from advanced imaging have outstripped all other physician service fees [8]. CT scan for flank pain may also trigger expensive work-ups for incidental findings, further contributing to inefficient, costly care [9].

Evidence is needed to guide CT imaging in patients with acute flank pain. Recently, a panel of decision rule experts identified atraumatic flank pain as one of the 10 highest priority clinical problems for the development of clinical decision rules [10]. A successful clinical decision rule for acute flank pain would help physicians identify which patients with acute flank pain benefit diagnostically from CT imaging, and conversely, identify which patients in whom CT may be avoided [11]. Investigators recently developed a clinical prediction rule for the identification of ureteral stone: the STONE score sorts patients with suspected ureterolithiasis into low-, moderate-, and high-risk groups, with those with a high score in the original study having an 89% probability of a ureteral stone and a 1.6% probability of an important alternative diagnosis [7]. On external validation, the STONE score successfully sorted patients into risk groups, but a *high* score had a sensitivity of only 53% for ureteral stone and the upper limit of the 95% confidence interval (CI) for the probability of an alternative diagnoses was 3.6% [6]. The STONE score was not specifically developed to exclude clinically important ureteral stones (i.e. ureteral stone with urosepsis) in patients with flank pain [11].

The long-term goal of this research is to reduce unnecessary CT imaging of patients presenting with acute flank pain by developing a successful clinical decision rule. Toward this goal, this is an exploratory study of participants presenting to the ED with acute flank pain, in which we determined the prevalence of clinically important outcomes in patients with acute flank pain and identified candidate clinical criteria for potential decision rules that predict these outcomes. These initial steps will allow for the planning of a large, prospective clinical decision rule derivation and validation study to safely reduce CT imaging in patients with acute flank pain, including the determination of the prevalence of clinically important outcomes, as well as the identification of important predictors.

2. Methods

2.1. Study design/setting

We performed this retrospective analysis using data from the Study of Ultrasonography versus Computed Tomography for Suspected Nephrolithiasis (trial registration number: NCT01451931 at clinicaltrials.gov) [5], a randomized comparative effectiveness trial that was conducted at 15 academic emergency departments across the United States between October 2011 and February 2013. We obtained institutional review board approval for this research from the Committee on Human Research.

2.2. Participants

In the parent study, adult patients who required imaging (as determined by an attending emergency physician) for acute flank pain suspicious were randomly assigned to receive point-of-care (POC) ultrasound, radiology ultrasound, or CT as their initial imaging test. Patients were excluded from enrollment if they were pregnant, at high risk of an important alternative (non-kidney stone) diagnosis (as determined by the ED provider), had received a kidney transplant, required dialysis, had a known solitary kidney, or if they were a male weighing >285 lb or female weighing >250 lb.

2.3. Measurements

Prior to patient enrollment, research coordinators, who were blinded to the study hypotheses, attended a two-day meeting to receive training regarding study protocol, forms, and data collection. They also participated in weekly online meetings to assure ongoing data collection consistency. Research coordinators used a standardized data collection form to collect detailed demographic, clinical, laboratory, and imaging data during the index ED. Patients were directly interviewed by research personnel for the subjective variables during the index ED visits (pain level, nausea, vomiting, time since onset of pain in hours, pain similar to prior stone, dysuria). All data were recorded on paper forms and faxed to a data-coordinating center, which provided immediate feedback for form completeness.

2.4. Outcomes

For this analysis, three emergency physicians and a radiologist (RCW, RR, JF, RSB) defined the main outcome as clinically important outcomes which required inpatient admission, including ureteral stones and non-stone diagnoses such as appendicitis, cholecystitis, pyelonephritis, and ovarian pathology requiring inpatient admission (chosen by consensus

from the alternative diagnoses identified in the parent randomized trial and prior literature) [5,12]. See Table 2 for the list of clinically important outcomes. These participants were all admitted as part of their management in the original trial. We defined our second outcome, "clinically important stone" - as ureteral stone requiring urologic intervention up to 30 days after the index emergency department visit (this cutoff was chosen because most trials and studies of observation for ureteral stone passage use this time point) [13–15]. Urologic interventions included ureteroscopy, lithostripsy, percutaneous nephrectomy, or stent placement. Regarding these outcomes, participants were interviewed during the baseline visit, were followed throughout hospitalization and then contacted over the ensuing 30 days to assess their occurrence. Research assistants also reviewed medical records for each participant at 30 days.

2.5. Predictor variables

The candidate predictor variables captured in the randomized trial are listed in Appendix 1. We reviewed prior studies of clinical decision rules and studies identifying predictors of ureteral stones requiring intervention or serious alternative diagnoses [7,16–19]. Important predictors of stone requiring urologic intervention from the literature review included stone size, stone location, pain level, signs of urinary tract infection (elevated white blood cell count, leukocyte esterase and nitrites on urinalysis), and age. We chose to exclude CT scan findings as candidate variables (the presence of ureteral stone, stone size and location) because our goal was to develop a decision rule to reduce CT use. We included hydronephrosis on imaging as a candidate variable, as hydronephrosis can be identified reliably and with moderate to excellent sensitivity on ultrasound [20,21]. Because not all ED clinicians are proficient at emergency ultrasound and because predictors obtained from routine history, physical exam, and laboratory tests may be the most simple to use and acceptable to clinicians, we chose to develop clinical decision rules both with and without the finding of hydronephrosis.

2.6. Statistical analysis

Prior to analyses, we delineated our target decision rule sensitivity to be 98%, consistent with other decision rules to identify serious outcomes. We developed 4 separate multivariate models - 2 to predict clinically important ureteral stones and 2 for the combined clinically important diagnoses. Because emergency physicians are concerned with both stone and non-stone diagnoses in patients with acute flank pain, we designated the combined clinically important outcomes as the primary outcome, and clinically important stone as a secondary outcome. We used χ^2 recursive partitioning to construct a decision tree to identify predictors to for both outcomes. χ^2 recursive partitioning was chosen as the modeling method (vs. logistic regression) because the objective was to derive a highly sensitive decision rule to exclude important outcomes. Recursive partitioning has been used to derive a number of well-known clinical decision rules, such as the NEXUS Cervical Spine and the PECARN head injury rules [22,23].

We used the rpart package in R (R Core Team [36]; R Foundation for Statistical Computing, Vienna, Austria), and included all variables as candidate predictors. A list of potential predictors, and how the predictors were coded can be found in Appendix 1. For continuous

variables (WBC count, hemoglobin, and age), we used k means clustering to choose cut-points (see Appendix 1) in order to improve accuracy and decrease over-fitting [24,25]. The outcomes were coded as binary outcomes. In order to generate a high sensitivity decision instrument, we specified a loss matrix of 5:1 to favor false negatives.

2.7. Missing data

The rpart program uses a native algorithm of "surrogate splits" to handle missing data in the predictor variables (when a value for a predictor variable is missing, and that variable needs to be used to determine a split, an alternative variable that is highly correlated with the missing variable is used to determine the direction of the split) [26]. Thus, we used the entire cohort for outcomes were recorded regardless of missing data among predictors, depending on the surrogate split function. The proportion of missing data is displayed in Appendix 1. Four patients out of 2759 (0.1%) were missing data related to admission, the primary outcome. All 2759 patients had outcomes recorded for the secondary outcome, urologic intervention within 30 days. Less than 2.5% of data was missing for all candidate predictors except for the following serum and urine studies, which had approximately 12–13% missing: WBC count, hemoglobin level, hematuria and pyuria on urinalysis. We did not include urine dipstick as a candidate variable, as there was 40% missing; 6 of 15 ED sites from the original trial did not routinely use urine dipstick testing. To determine whether test characteristics resulting from rpart classification were sensitive to its use of surrogate splits used on missing data, we compared the sensitivity and specificity to that of the corresponding model fitted to the subset with complete data on all predictors (Appendix 2).

A second sensitivity analysis was conducted to compare the decision rules for clinically important stone including hydronephrosis as a potential predictor. We compared the decision rule derived on the entire dataset to a decision rule derived on a cohort who received ultrasound as the index test, and those who received CT were removed (N = 1733). This was performed to determine whether the test characteristics were sensitive to the imaging modality to identify hydronephrosis (Appendix 3).

A final secondary analysis was included to determine if the decision rule for clinically important outcomes would differ depending on whether the outcome was defined as patients requiring admission at the index visit, if we included those admitted up to day 7 after the index visit. Thus we identified subjects admitted to the hospital after the initial admission, up to day 7. An additional 47 subjects were identified. Recursive partitioning was used to construct a potential decision rule (Appendix 4), and test characteristics were reported in Appendix 3.

3. Results

Of the 2759 participants, the median age was 40, 1428 (51.7%) were male, and 1128 (40.9%) were White. Additional characteristics of the participants are described in Table 1. 236 (8.6%) participants admitted to the hospital and thus were considered to have the primary outcome. Of those admitted, 131 (4.9%) patients were admitted for an important alternative diagnosis, such as appendectomy, laparotomy or laparoscopic repair of ovarian torsion, cholecystectomy, or biopsy of a suspicious mass. Table 2 displays the list of

clinically important outcomes in admitted participants in descending frequency. An additional 47 subjects were admitted to the hospital after their index ED visit, up to 7 days. 143/2759 (5.2%) of participants required a urologic intervention by 30 days after the index ED visit, and were considered to have the secondary outcome. Fifty-two (1.9%) participants received a urologic intervention during the index visit.

Fig. 1a shows a decision tree constructed to predict clinically important outcomes. This figure shows the predictor variables chosen by the recursive partitioning applied to the entire cohort until a low risk group with very few cases remains. Predictors of clinically important outcomes include anemia (hemoglobin <13.2 g/dl), WBC count >11 000/μl, age > 42 years, and the absence of CVAT. Participants with the absence of any predictor are at low risk of the outcome, with a prevalence of clinically important stone = 1.1% (95% CI 0.3–2.4%). Fig. 1b shows a decision tree to predict clinically important outcomes in which hydronephrosis was included as a predictor variable. In this model, predictors of clinically important outcomes include WBC count >11 000/μl, age > 42 years, duration of symptoms >12 h. Participants with the absence of any predictor are at low risk with a prevalence of clinically important outcomes = 1.4% (95% CI 0.6–2.9%). Hydronephrosis was not an important predictor of clinically important outcome requiring admission.

Fig. 2a shows a decision tree to predict clinically important stone. Predictors of clinically important stone include a prior history of stone, nausea, and maximal pain level of 10/10. Participants with the absence of any predictor had a prevalence of clinically important stone = 0.2% (95% CI 0–1.1%). Fig. 2b shows a decision tree constructed including hydronephrosis as a candidate predictor variable. Predictors of clinically important stone include the presence of hydronephrosis, a prior history of stone, and WBC count ≥8400/μl. Participants with the absence of any predictor had a prevalence of clinically important stone = 0.3% (95% CI 0–1.1%).

Table 3 presents each potential decision rule's respective classification performance. The potential decision instrument for clinically important outcomes has a sensitivity of 97.9% (95% CI: 94.8–99.2%) and the specificity of 18.7% (95% CI: 17.2–20.2%), and a negative likelihood ratio of 0.11 (95% CI: 0.05–0.27). The failure rate, or proportion of subjects positive for the outcome that the decision rule identified as negative, was 4/475, or 1.1% The 5 cases in which the decision rule failed were the following final hospital diagnoses: two cases of urolithiasis, one which required intervention during the initial hospitalization, one case of appendicitis, one case of pyelonephritis, and one case of suspected cancer. The addition of hydronephrosis to the available predictors did not improve the sensitivity or specificity. The potential decision instrument for clinically important stone has a sensitivity of 99.3% (95% CI: 95.5–100%) and a specificity of 18.1% (95% CI: 16.6–19.6%), and a negative likelihood ratio of 0.04 (95% CI: 0.01–0.27). The addition of hydronephrosis as a candidate predictor resulted in a decision rule with a similar sensitivity, negative predictive value and negative likelihood ratio 0.05 (95% CI 0.01–0.24), but significantly higher specificity - 26.0% (95% CI 24.2–27.7%). The failure rates of the clinically important stone decision rules were 1/474 (0.2%) and 2/671 (0.3%), respectively.

Appendix 2 presents a sensitivity analysis for the robustness of the four models with alternative treatment of missing data. We compared test characteristics of decision trees derived from the full dataset (using rpart and its native surrogate splits algorithm to classify all observations) to those of decision trees derived from a subset of the data that excluded observations with missing predictor data. The resulting classification trees produced the same variables, with similar cut points. Overall, the models performed similarly in the complete data, with the exception of the clinically important stone -hydronephrosis rule, which had a significantly lower specificity (26% vs. 16%), likely because this rule included WBC as a predictor, which had significant missing values. Otherwise, the results do not appear sensitive to the missing data in the predictors, as the potential decision instruments are identical, with similar test characteristics.

Appendix 3 also shows the sensitivity analysis in which those who received CT scan were removed – an ultrasound only cohort. The decision rule that was derived only consisted of 2 variables, hydronephrosis, and nausea. The low risk group (no hydronephrosis, no nausea) had a similar, or even superior test characteristics compared to the decision rule derived on the entire cohort, including those who received CT scan. Appendix 3 also shows the test characteristics of a decision rule for clinically important outcomes, when the outcome includes those admitted up to 7 days after the index visit. The test characteristics are similar to those of the decision rule for those admitted at the index visit. Appendix 4 is a figure, displaying this additional decision rule model.

4. Limitations

The primary limitations of this study arise from the retrospective design with data derived from a clinical trial, which led to high rates of missing candidate criteria. Optimal clinical decision rule development utilizes prospective candidate criteria assessment with data collection forms designed specifically for the purpose of decision rule development [11,27]. Other variables (abnormal vital signs, urine nitrites, leukocyte esterase, and serum creatinine level) that were not captured in the parent trial are potentially strong predictors of clinically important outcomes and to lesser degree, clinically important stones. Their inclusion could produce decision instruments with potentially improved sensitivity and specificity. We plan to capture these variables in future prospective studies. We conducted an analysis to determine whether the results are sensitive to missing predictors by applying our decision rules to only the subset of the cohort that had complete predictor data, which resulted in nearly identical test performance. Other limitations of this retrospective analysis include the lack of assessment of inter-rater reliability of candidate criteria.

Aside from sensitivity and specificity, another means to assess the value of a clinical prediction rule is to weigh the miss rate (proportion of subjects in which the outcome was present and the decision rule identified the patient as low risk) and the potential improvement in efficiency (proportion of subjects in which the test was negative/the entire cohort). Successful clinical decision rules should have a low miss rate and substantial improvements in efficiency. In the decision rule for clinically important outcomes, the reduction in CT ordering of 17% vs. miss rate of 1.1% would suggest that this decision rule needs additional refinement.

The long-term goal is to develop a decision instrument to evaluate all patients with acute, atraumatic flank pain in which ureterolithiasis is suspected. However, the cohort of patients used for this study is not precisely representative of the patients on which a clinical decision rule would be applied. Patients at high risk for alternative diagnoses and certain stone related emergencies were excluded from the randomized trial, as were those who clinicians did not intend to image. Despite these limitations, our cohort is similar to those in previously published reports; approximately 5% of subjects enrolled in the randomized trial were suspected of an alternative diagnosis [7] Also, the rate of admission to the hospital was approximately 9%, which is similar to the 10% hospitalization rate reported by prior reports using the National Hospital Ambulatory Care Survey [1,4].

5. Discussion

In this exploratory study, we identified participants who had clinically important outcomes, which we defined as inpatient admission for ureteral stone or an important alternative diagnosis. We found that approximately 9% of participants in the multi-center trial had a clinically important outcome, and 5% had a clinically important stone. Using recursive partitioning, we derived a potential decision rule for clinically important outcomes, which is highly sensitive, with an excellent negative predictive value and a negative likelihood ratio of approaching 0.1. The specificity of the decision rule for clinically important outcomes is disappointing, likely due to missing variables such as vital sign abnormalities, and creatinine level. The addition of hydronephrosis did not improve the accuracy of the decision rule, likely because hydronephrosis is a strong predictor of ureteral stone requiring intervention but not of alternative diagnoses. Nonetheless, we identified important predictors of clinically important outcomes, such as increasing age, the absence of CVAT, elevated WBC, and anemia. These predictors should be measured in similar efforts in the future. The 2 preliminary decision rules for clinically important stones have excellent sensitivities and negative predictive values with clinically useful negative likelihood ratios of less than 0.1. The decision rule incorporating hydronephrosis exhibits significantly higher specificity: 26.0% (95% CI 24.2–27.7%) vs. 18.1% (95% CI: 16.6–19.6%), and could ultimately result in a higher proportion of patients being identified as low risk using ultrasound and thereby spared CT. If such a rule was validated with similar test characteristics, approximately a quarter of CT scans could be avoided while missing 0.3% (95% CI 0.04–1.1%) clinically important stones. Important predictors to be considered would be prior history of stone, the presence of hydronephrosis, 10/10 pain level, nausea, and elevated WBC count.

This study differs from other CDRs or predictor-finding studies for acute flank pain as it seeks to explicitly address 2 important clinical outcomes in patients who present to the ED with acute flank pain without information from CT scan. First, we derived a high sensitivity decision rule for clinically important outcomes – a combined outcome of ureteral stone and non-stone alternative diagnoses that require admission. This is the first study to identify clinical predictors of a combined stone and non-stone outcomes, which we believe is conceptually important as emergency physicians order CT scan to identify both stones that require management as well as non-stone alternative diagnoses [12,28–30]. Other studies have predicted the need for urologic intervention, but require information from CT scan, and thus cannot be used to avoid CT [17,18]. Our results confirm findings from prior studies

using urologic intervention as the outcome. The absence of hydronephrosis on ultrasound has been reported to predict low rates of urologic intervention among those with suspected stone [16,17]. The finding of hydronephrosis on renal point-of-care limited ultrasonography was shown to have a sensitivity of 66% and specificity of 58% for urologic intervention; moderate to severe hydronephrosis had a modest specificity (86%), but the sensitivity was diminished (36%). The addition of renal point-of-care limited ultrasonography modestly improved risk stratification of the STONE score [31]. Age and elevated white blood cell count are known predictors of ureteral stone requiring urologic intervention [17,32]. A prior history of kidney stone is known to increase the risk of ureteral stone in patients with suspected kidney stone [33]. By combining several important predictors using recursive partitioning, we developed a multivariable test with a near perfect sensitivity and acceptable specificity.

These decision instruments are not ready for clinical use. While we have shown that it is feasible to derive decision rules for acute flank pain, all of these decision instruments require further refinement and validation. However, we believe that this exploratory study provides a conceptual blueprint to develop a successful CDR for acute flank pain. Similar to prior studies of successful decision rule development, such as the PECARN head injury rule, we selected the study outcomes by focusing on clinical outcomes in patients with acute flank pain who require intervention or inpatient treatment [28, 30,34]. We used recursive partitioning to derive a decision instrument with a high sensitivity, could exclude clinically important diagnoses at the bedside, similar to the PECARN head injury rule [27,35], potentially allowing clinicians to avoid CT if validated. In order to validate a similar decision rule with a desired sensitivity for clinically important outcomes of 98% or greater (with a 95% confidence interval width of 2% [96–100%]), approximately 2100 participants would need to be enrolled.

In conclusion, we have determined the prevalence of clinically important outcomes and derived preliminary clinical decision rules to guide selective imaging in patients presenting with acute flank pain to the ED. These results should inform future prospective studies to derive and validate such rules.

Acknowledgments

Funding/support

This study was supported by funding from the Agency of Healthcare Research and Quality Grant K08 HS02181 (Wang) and K24 CA125036 (Smith-Bindman).

We would like to acknowledge Dr. Jane Hall for her contribution to the recursive partitioning data analysis.

References

1. Foster, G., Stocks, C., Borofsky, MS. Statistical brief #139. Jul 27. 2012 p. 1-10.
2. Fwu C-W, Eggers PW, Kimmel PL, Kusek JW, Kirkali Z. Emergency department visits, use of imaging, and drugs for urolithiasis have increased in the United States. Kidney Int. Apr; 2013 83(3): 479–86. [PubMed: 23283137]
3. Hyams ES, Korley FK, Pham JC, Matlaga BR. Trends in imaging use during the emergency department evaluation of flank pain. J Urol. Dec; 2011 186(6):2270–4. [PubMed: 22014815]

4. 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. Jul 15; 2011 18(7):699–707. [PubMed: 21762233]
5. Smith-Bindman R, Aubin C, Bailitz J, et al. Ultrasonography versus computed tomography for suspected nephrolithiasis. N Engl J Med. Sep 18; 2014 371(12):1100–10. [PubMed: 25229916]
6. Wang RC, RR 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. 2015 Accepted for Publication.
7. 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. Apr 26.2014 348(mar26 2):g2191. [PubMed: 24671981]
8. Smith-Bindman R, Miglioretti DL, Larson EB. Rising use of diagnostic medical imaging in a large integrated health system. Health Aff. Jan 1; 2008 27(6):1491–502.
9. Thompson RJ, Wojcik SM, Grant WD, Ko PY. Incidental findings on CT scans in the emergency department. Emerg Med Int. 2011; 2011:624847. [PubMed: 22046542]
10. Finnerty NM, Rodriguez RM, Carpenter CR, et al. Clinical decision rules for diagnostic imaging in the emergency department: a research agenda. Acad Emerg Med. Dec; 2015 22(12):1406–16. [PubMed: 26567885]
11. Green SM, Schriger DL. The sinking STONE: what a failed validation can teach us about clinical decision rules. Ann Emerg Med. Jan 21.2016
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 Off J Soc Acad Emerg Med. Jun 14; 2013 20(5):470–8.
13. Pickard R, Starr K, MacLennan G, Lam T, Thomas R. Medical expulsive therapy in adults with ureteric colic: a multicentre, randomised, placebo-controlled trial. Lancet. 2015
14. Vincendeau S, Bellissant E. Tamsulosin hydrochloride vs placebo for management of distal ureteral stones: A multicentric, randomized, double-blind trial. Archives of …. 2010
15. Pearle MS, Calhoun EA, Curhan GC. Project UDoA. Urologic diseases in America project: urolithiasis. J Urol. Mar 1; 2005 173(3):848–57. [PubMed: 15711292]
16. Yan JW, McLeod SL, Edmonds ML, Sedran RJ, Theakston KD. Normal renal sonogram identifies renal colic patients at low risk for urologic intervention: a prospective cohort study. Cjem. Jan; 2015 17(1):38–45. [PubMed: 25781382]
17. Yan JW, McLeod SL, Edmonds ML, Sedran RJ, Theakston KD. Risk factors associated with urologic intervention in emergency department patients with suspected renal colic. J Emerg Med. Aug; 2015 49(2):130–5. [PubMed: 25910829]
18. Papa L, Stiell IG, Wells GA, Ball I, Battram E, Mahoney JE. Predicting intervention in renal colic patients after emergency department evaluation. Cjem. Mar 1; 2005 7(2):78–86. [PubMed: 17355656]
19. Schoenfeld EM, Poronsky KE, Elia TR, Budhram GR, Garb JL, Mader TJ. Young patients with suspected uncomplicated renal colic are unlikely to have dangerous alternative diagnoses or need emergent intervention. West J Emerg Med. Mar; 2015 16(2):269–75. [PubMed: 25834669]
20. Herbst MK, Rosenberg G, Daniels B, et al. Effect of provider experience on clinician-performed ultrasonography for hydronephrosis in patients with suspected renal colic. Ann Emerg Med. Sep; 2014 64(3):269–76. [PubMed: 24630203]
21. Sternberg KM, Pais VM Jr, Larson T, Han J, Hernandez N, Eisner B. Is hydronephrosis on ultrasound predictive of ureterolithiasis in patients with renal colic? J Urol. May 3.2016
22. Kuppermann N, Holmes JF, Dayan PS, et al. Identification of children at very low risk of clinically-important brain injuries after head trauma: a prospective cohort study. Lancet. 2009; 374(9696):1160–70. [PubMed: 19758692]
23. Stiell IG, Clement CM, McKnight RD, et al. The Canadian C-spine rule versus the NEXUS low-risk criteria in patients with trauma. N Engl J Med. Dec 25; 2003 349(26):2510–8. [PubMed: 14695411]

24. Peng Y, Zhang Y, Kou G, Shi Y. A multicriteria decision making approach for estimating the number of clusters in a data set. PLoS One. 2012; 7(7):e41713. [PubMed: 22870181]
25. Rouzbahman M, Jovicic A, Chignell M. Can cluster-boosted regression improve prediction: death and length of stay in the ICU? IEEE J Biomed Health Inform. Feb 3.2016
26. Tierney NJ, Harden FA, Harden MJ, Mengersen KL. Using decision trees to understand structure in missing data. BMJ Open. 2015; 5(6):e007450.
27. Stiell IG, Wells GA. Methodologic standards for the development of clinical decision rules in emergency medicine. Ann Emerg Med. May 01; 1999 33(4):437–47. [PubMed: 10092723]
28. Wang RC. Managing urolithiasis. Ann Emerg Med. Nov 23.2015
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. JURO. 2006; 175(5):1725–30.
30. Teichman JMH. Clinical practice. Acute renal colic from ureteral calculus. N Engl J Med. Feb 12; 2004 350(7):684–93. [PubMed: 14960744]
31. Daniels B, Gross CP, Molinaro A, et al. STONE PLUS: evaluation of emergency department patients with suspected renal colic, using a clinical prediction tool combined with point-of-care limited ultrasonography. Ann Emerg Med. Apr; 2016 67(4):439–48. [PubMed: 26747219]
32. Schoenfeld E, Poronsky K, Elia T, Budhram G, Garb J, Mader T. Young patients with suspected uncomplicated renal colic are unlikely to have dangerous alternative diagnoses or need emergent intervention. West J Emerg Med. Apr 23; 2015 16(2):269–75. [PubMed: 25834669]
33. Goldstone A, Bushnell A. Does diagnosis change as a result of repeat renal colic computed tomography scan in patients with a history of kidney stones? Am J Emerg Med. Apr; 2010 28(3): 291–5. [PubMed: 20223385]
34. Ha M, MacDonald RD. Impact of CT scan in patients with first episode of suspected nephrolithiasis. J Emerg Med. Oct; 2004 27(3):225–31. [PubMed: 15388206]
35. Green SM, Schriger DL, Yealy DM. Methodologic standards for interpreting clinical decision rules in emergency medicine: 2014 update. Ann Emerg Med. Sep; 2014 64(3):286–91. [PubMed: 24530108]
36. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013. http://www.R-project.org/

Appendix 1 List of candidate predictors

Candidate predictors	Missing observations (%)	Coding
Gender	0	Male = 1
Age	0	18–24 yrs
		25–32 yrs
		33–41 yrs
		42–52 yrs
		53–81 yrs
Race	16 (0.6)	White
		African American
		Asian
		Native American
		Pacific Islander
		More than one
		Hispanic
		Missing

Candidate predictors	Missing observations (%)	Coding
Duration of pain since onset	23 (0.8)	1–2 h
		3–6 h
		7–12 h
		13–24 h
		25–48 h
		>48 h
		Refused
Pain level	2 (0.1)	1–10
Abdominal guarding	58 (2.1)	Yes = 1, Voluntary = 2
Murphy's sign	80 (2.9)	Yes = 1
RLQ tenderness	31 (1.2)	Yes= 1
LLQ tenderness	33 (1.2)	Yes = 1
Nausea	8 (0.3)	Yes = 1
Vomiting	12 (0.4)	Yes = 1
Dysuria	32 (1.2)	Yes = 1
Prior kidney stone	58 (2.1)	Yes = 1
Prior urologic intervention	72 (2.6)	Yes = 1
Pain similar to prior stone	110 (4.0)	Yes = 1
Prior history of cancer	11 (0.4)	Yes = 1
CVAT, any	45 (1.6)	Yes = 1
Hematuria on urinalysis	339 (12.6)	<3 rbc/hpf
		>3 rbc/hpf
		TNTC
Urine WBC	351 (12.8)	<50 wbc/hpf
		>50 wbc/hpf
		TNTC
White blood count	365 (13.3)	2.2–6 thousands/μl
		6.1–8.3 thousands/μl
		8.4–11 thousands/μl
		11.1–14.7 thousands/μl
		14.8–29.7 thousands/μl
Hemoglobin (sd)	337 (12.1)	3.9–11.2 g/dl
		11.3–13.2 g/dl
		13.3–14.6 g/dl
		14.7–15.9 g/dl
		16–19.8 g/dl
Hydronephrosis/hydroureter	60 (2.2)	Yes = 1

Appendix 2 Test characteristics of decision instruments for acute flank pain in complete data

	95% CI					
	Sensitivity	**Specificity**	**Negative predictive value**	**Positive predictive value**	**Negative likelihood ratio**	**Positive likelihood ratio**
Clinically important outcome (Flank pain requiring admission, prevalence = 9.9%)						
TP: 231	97.8%	16.0%	98.5%	8.5%	0.14	1.16
TN: 329	(94.6–99.2%)	(14.4–17.6%)	(96.3–99.4%)	(8.4–8.7%)	(0.06–0.33)	(1.13–1.20)
FP: 1733						
FN: 5						
Clinically important outcome (Hydronephrosis included as a predictor)						
TP: 229	97.0%	19.1%	98.6%	10.1%	0.15	1.2
TN: 482	(93.7–98.7%)	(17.6–20.7%)	(96.9–99.4%)	(8.9–11.4%)	(0.07–0.32)	(1.16–1.23)
FP: 2037						
FN: 7						
Clinically important stone (Requiring urologic intervention, prevalence = 5.2%)						
TP: 140	99.3%	18.4%	99.8%	6.3%	0.04	1.2
TN: 473	(95.5–100%)	(16.9–20.0%)	(98.6–100%)	(5.3–7.4%)	(0.01–0.27)	(1.18–1.24)
FP: 2099						
FN: 1						
Clinically important stone (Hydronephrosis included as predictor)						
TP: 137	100%	19.5%	100%	6.4%	0.00	1.24
TN: 485	97.3–100%	18.0–21.1%	99.0–100%	5.4–7.5%	(0.01–NA)	(1.22–1.27)
FP: 2003						
FN: 0						

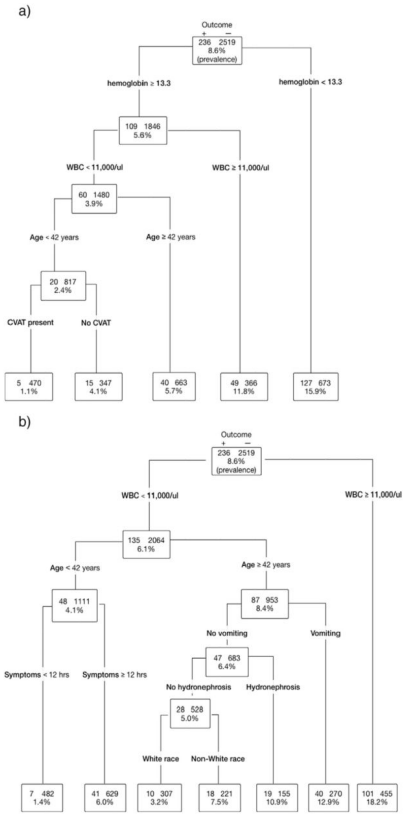
TP = true positive; TN = true negative; FP = false positive; FN = false negative.

Appendix 3 Test characteristics of decision instruments for acute flank pain, sensitivity analyses

	95% CI					
	Sensitivity	**Specificity**	**Negative predictive value**	**Positive predictive value**	**Negative likelihood ratio**	**Positive likelihood ratio**
Clinically important stone in the ultrasound only cohort (n = 1733)						
TP: 81	98.8%	28.2%	99.8%	6.4%	0.04	1.4
TN: 466	(92.5–99.9%)	(26.1–30.5%)	(98.6–100%)	(5.1–7.9%)	(0.01–0.30)	(1.32–1.43)
FP: 1185						
FN: 1						
Clinically important outcomes, all admitted patients up to day 7						
TP: 276	97.5%	16.5%	98.3%	11.8%	0.15	1.2
TN: 407	(94.8–98.9%)	(15.0–18.0%)	(96.4–99.3%)	(10.5–13.2%)	(0.07–0.31)	(1.14–1.20)
FP: 2056						
FN: 7						

TP = true positive; TN = true negative; FP = false positive; FN = false negative.

Appendix



Appendix 4.
Decision tree for clinically important outcome admission up to day 7

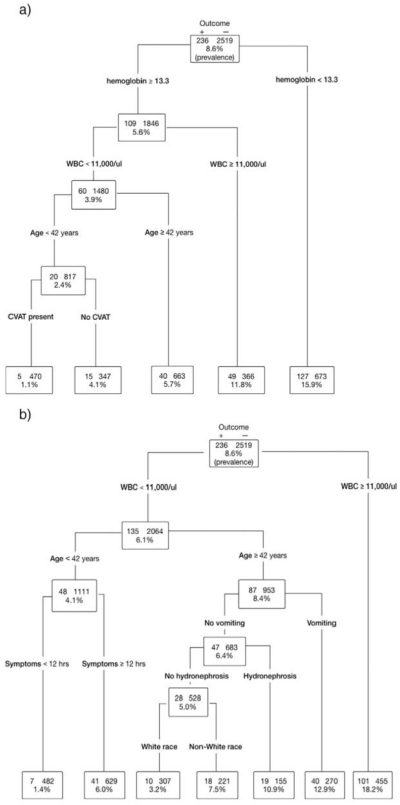


Fig. 1.

a. Decision tree for clinically important outcome. b. Decision tree for clinically important outcome – hydronephrosis.

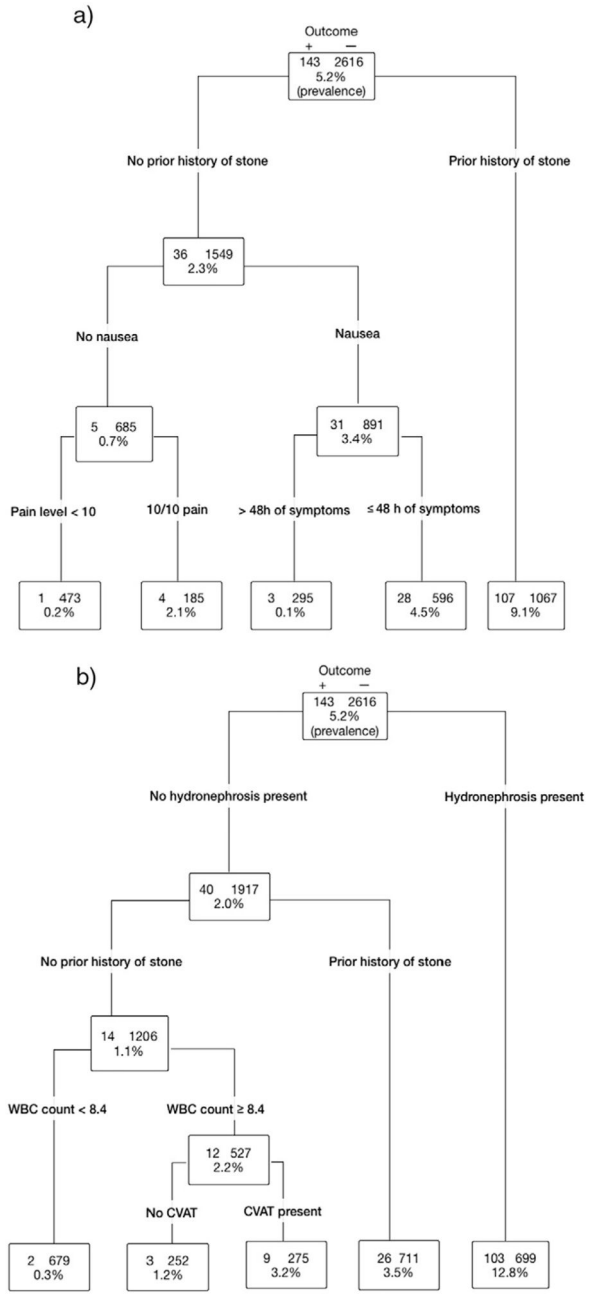


Fig. 2.

a, b. Decision tree for clinically important stone.

Table 1

Baseline characteristics of the 2759 participants.

Median age (IQR)	40 (30–50)
Male	1428 (51.8)
Race	
White	1128 (40.9)
African American	690 (25.0)
Asian	125 (4.5)
Native American	38 (1.4)
Pacific Islander	6 (0.2)
More than one	88 (3.2)
Hispanic	668 (24.2)
Refused	16 (0.6)
Median pain level (IQR)	9 (7–10)
Duration of pain since onset (hours)	
1–2	445 (16.1)
3–6	465 (16.9)
7–12	270 (9.8)
13–24	284 (10.3)
25–48	292 (10.6)
>48	980 (35.5)
Refused	23 (0.8)
Nausea	1750 (63.4)
Prior diagnosis of kidney stone	1149 (41.7)
Prior urologic intervention	375 (13.6)
Costo-vertebral angle tenderness	1448 (52.4)
Hematuria on urinalysis	
<3 rbc/hpf	949 (34.4)
>3 rbc/hpf	1215 (44.0)
Too numerous to count	256 (9.3)
Not obtained	339 (12.2)
WBC on urinalysis	
<50 wbc/hpf	2160 (78.3)
>50 wbc/hpf	192 (6.9)
Too numerous to count	56 (2.0)
Not obtained	351 (12.7)
The presence of hydronephrosis on imaging	
None	1897 (68.8)
Present	802 (29.1)
Not reported	58 (2.1)
Admitted to hospital	236 (8.6)
Urologic intervention	

Received urologic intervention at baseline	52 (1.9)
Received urologic intervention by 30 days	143 (5.2)

Table 2

List and frequency of clinically important outcomes requiring admission, N = 236.

Urolithiasis requiring admission	105
Pyelonephritis/UTI	34
Cancer evaluation	14
Appendicitis	11
Diverticulitis/colitis	10
Symptomatic cholelithiasis/cholecystitis	9
Non-specific pain	6
Pancreatitis	5
Pneumonia/pleural effusion	5
Musculoskeletal	5
Cardiovascular	4
Peptic ulcer disease/non-specific vomiting	4
Testicular/ovarian torsion	4
Genitourinary abnormality (i.e. ureterocele)	4
Intra-abdominal abscess	4
Soft tissue infection/hematoma	3
STD/PID	2
Pulmonary embolism, deep vein thrombosis	2
Kidney disease	2
Hepatitis/portal hypertension	1
Small bowel obstruction	1
Diabetic keto-acidosis	1

Table 3

Test characteristics of decision instruments for acute flank pain.

	95% CI					
	Sensitivity	**Specificity**	**Negative predictive value**	**Positive predictive value**	**Negative likelihood ratio**	**Positive likelihood ratio**
Clinically important outcome (Flank pain requiring admission) prevalence = 11.0%						
TP: 231	97.9%	18.7%	98.9%	10.1%	0.11	1.2
TN: 470	(94.8–99.2%)	(17.2–20.2%)	(97.4–99.6%)	(8.9–11.5%)	(0.05–0.27)	(1.17–1.24)
FP: 2049						
FN: 5						
Clinically important outcome; hydronephrosis included as predictor						
TP: 229	97.0%	19.1%	98.6%	10.1%	0.15	1.2
TN: 482	(93.7–98.7%)	(17.6–20.7%)	(96.9–99.4%)	(8.9–11.4%)	(0.07–0.32)	(1.16–1.23)
FP: 2037						
FN: 7						
Clinically important stone (Requiring urologic intervention) prevalence = 5.2%						
TP: 142	99.3%	18.1%	99.8%	6.2%	0.04	1.2
TN: 473	(95.5–100%)	(16.6–19.6%)	(98.6–100%)	(5.2–7.3%)	(0.01–0.27)	(1.18–1.24)
FP: 2143						
FN: 1						
Clinically important stone; hydronephrosis included as predictor						
TP: 141	98.6%	26.0%	99.7%	6.8%	0.05	1.3
TN: 679	94.5–99.7%	24.2–27.7%	98.8–99.9%	5.8–7.9%	(0.01–0.24)	(1.3–1.4)
FP: 1937						
FN: 2						

TP = true positive; TN = true negative; FP = false positive; FN = false negative.