Journal of HOSPITAL MEDICINE www.journalofhospitalmedicine.com

ORIGINAL RESEARCH

External Validation of Risk Stratification Strategy in the Use of Renal Ultrasonography in the Evaluation of Acute Kidney Injury

Ivan K. Ip, MD, MPH[1,2,3]*, Patricia C. Silveira, MD[1,2], Emily C. Alper, BA[1,2], Carol B. Benson, MD[2], Ramin Khorasani, MD, MPH[1,2]

[1]Center for Evidence-Based Imaging, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; [2]Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; [3]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts.

BACKGROUND: Per the American College of Radiology Appropriateness Criteria, renal ultrasound is the most appropriate imaging examination to evaluate patients with acute kidney injury. However, recent studies suggest that renal ultrasound may be more selectively performed, which could lead to reductions in the use of medical imaging.

OBJECTIVE: Evaluate a published risk stratification prediction model (the Licurse model) for using renal ultrasound in hospitalized patients with acute kidney injury.

DESIGN: Prospective, observational cohort study.

SETTING: A 793-bed, quaternary care, academic hospital.

PATIENTS: All adult hospitalized patients who underwent renal ultrasound for the indication of acute kidney injury.

INTERVENTION/EXPOSURE: None.

MEASUREMENTS: Primary outcome was rate of hydronephrosis diagnosed on ultrasound. Secondary outcome was rate of hydronephrosis resulting in urologic intervention.

RESULTS: Of 778 patients who underwent renal ultrasonography to evaluate acute kidney injury, hydronephrosis was present in 106 (13.6%); urologic intervention was performed in 23 patients (3.0%). The Licurse model had sensitivity of 91.3% (95% confidence interval [CI]: 73.2%-97.6%) for urologic intervention and 93.4% (95% CI: 87.2%-96.8%) for hydronephrosis, respectively. Specificity was low for urologic intervention (23.0% [95% CI: 20.2-26.2]) and hydronephrosis (25.1% [95% CI: 22.0-28.6]). We estimated that for 22.6% of patients, hydronephrosis could be ruled out based on clinical predictors.

CONCLUSIONS: We found that the Licurse renal ultrasonography risk stratification model was sufficiently accurate in classifying patients at risk for ureteral obstruction among hospitalized patients with acute kidney injury. *Journal of Hospital Medicine* 2016;11:763–767. © 2016 Society of Hospital Medicine

According to the American College of Radiology Appropriateness Criteria, renal ultrasound (RUS) is the most appropriate imaging examination for evaluating patients with acute kidney injury (AKI), with a rating score of 9, representing the strongest level of recommendation.[1,2] However, recent studies suggest that RUS may be performed in patients with certain risk factors for ureteral obstruction,[1] which would lead to important reductions in the use of medical imaging. Licurse developed a risk stratification framework to help clinicians identify patients in whom RUS was most likely to be beneficial.[2] The model was built based on clinical predictors that included race, recent exposure to inpatient nephrotoxic medications, history of hydronephrosis, recurrent urinary tract infections, benign prostatic hyperplasia, abdominal or pelvic cancer, neurogenic bladder, single functional kidney, previous pelvic surgery, congestive heart failure, and pre–renal AKI. It was found, using a cross-sectional study design that included derivation and validation samples, that a low-risk population could be identified based on demographic and clinical risk factors; in this population, the prevalence of hydronephrosis, as well as the rate of hydronephrosis requiring an intervention, was only <1%.

However, due to several study limitations, including that it was performed at a single center,[3] the stratification prediction rule has yet to be adopted broadly. Although at least 1 other study has similarly found that RUS may not be efficacious in patients with no suggestive history and with other more likely causes for renal failure,[1] to the best of our knowledge, no large, external, prospective trial to validate the selective use of RUS in patients with AKI has been reported. Therefore, the aim of this study was to evaluate the accuracy and usefulness of the Licurse renal ultrasonography risk stratification model for hospitalized patients with AKI.

METHODS

Study Setting

The study site was a 793-bed academic, quaternary care, adult hospital with an affiliated cancer center.

*Address for correspondence and reprint requests: Ivan K. Ip, MD, Center for Evidence-Based Imaging, Department of Radiology and Medicine, Brigham and Women's Hospital, 20 Kent Street, 2nd Floor, Boston, MA 02445; Telephone: 617-525-9713; Fax: 617-525-7575; E-mail: iip@partners.org

Additional Supporting Information may be found in the online version of this article.

Received: December 3, 2015; **Revised:** March 24, 2016; **Accepted:** March 31, 2016
2016 Society of Hospital Medicine DOI 10.1002/jhm.2598
Published online in Wiley Online Library (Wileyonlinelibrary.com).

The requirement to obtain informed consent was waived by the institutional review board for this Health Insurance Portability and Accountability Act–compliant, prospective cohort study.

Study Population

The study cohort included all adult hospitalized patients who underwent an RUS for the indication of AKI over a 23-month study period, from January 2013 to November 2014. AKI was defined as having a peak rise in serum creatinine level of at least 0.3 mg/dL from baseline, based on data within the electronic health record (EHR). To ensure that the imaging study was not ordered for the purpose of follow-up or other reasons, patients who were renal transplant recipients, those who had ureteral stent or nephrostomy in place, patients who were recently diagnosed with hydronephrosis on prior imaging, and women who were pregnant were excluded based on retrospective chart review. In patients with multiple renal ultrasounds during the study period, only the first examination was considered.

Data Collection

We collected patient demographics in the study cohort from the EHR. Imaging data were identified using the radiology information system and computerized physician order entry (CPOE) system. For each eligible patient, we collected relevant clinical attributes including: (1) race, (2) history of hydronephrosis, (3) history of recurrent urinary tract infections, (4) history of benign prostatic hyperplasia, (5) history of abdominal or pelvic cancer, (6) history of neurogenic bladder, (7) history of single functional kidney, (8) history of previous pelvic surgery, (9) recent exposure to inpatient nephrotoxic medications, (10) history of congestive heart failure, and (11) history of pre–renal AKI. Information was collected from ordering clinicians at the time of imaging order entry using a computerized data capture tool integrated with the CPOE system. The data capture screen is shown in Supporting Figure 1 in the online version of this article. To validate the accuracy and completeness of this data entry, we manually reviewed objective clinical data from a random sample of 80 medical records for 480 clinical attributes. This number was selected based on a calculation of 80% power, 0.05 α, and a 0.1 proportion difference.

Patients received +1 point for the presence/absence of each clinical attribute. The sum of points was used to classify the patient's pretest probability of AKI as low (<2), medium (3), or high (>3). Both ordering and interpreting clinicians were blinded to the patient's prediction score.

Each RUS report was manually classified (by an internal medicine attending physician and a radiology trainee) as positive or negative for hydronephrosis, defined as any dilatation of the renal pelvis or the calyces. Subsequent use of urologic intervention was determined by full chart review of the sonographic positive cases. We defined these urologic interventions to include stent placement and nephrostomy tube placement. Only interventions performed during the same hospitalization as the index ultrasound were counted.

Outcomes

Our primary outcome was hydronephrosis (HN) diagnosed on ultrasound. Secondary outcome was hydronephrosis resulting in intervention (HNRI), defined as the need for urologic interventions of stent placement or nephrostomy tube placement.

Statistical Analysis

Analyses were performed using Microsoft Excel 2003 (Microsoft Corp., Redmond, WA) and JMP 10 (SAS Institute, Cary, NC). We used χ^2 to assess for differences in the rates of HN and HNRI across the 3 pretest probability risk groups. Sensitivity, specificity, negative predictive value, efficiency, and the number needed to screen to find 1 case of HN or HNRI for each risk group were calculated. The high and medium risk groups were merged for the purpose of calculating sensitivity and specificity. Efficiency was defined as the percentage of ultrasounds that could have been avoided based on applying the risk stratification model. We additionally performed a sensitivity analysis to evaluate how different cutoff thresholds for classifying low risk patients would affect the accuracy of the Licurse model. A 2-tailed P value of <0.05 was defined as statistically significant.

RESULTS

During the 23-month study period, a total of 961 RUS studies were completed for inpatients with AKI; 778 unique studies met our inclusion criteria (Figure 1).

Based on the manual review of objective clinical data from the random sample of 80 medical records for 480 clinical attributes, overall, there was 90.2% (433/480) concordance rate between the structured data entry and that captured in free text in the clinical notes. There were some variations in the concordance rates for each clinical attribute, ranging from 78.8% (63/80) for exposure to nephrotoxic drugs to 95% for history of congestive heart failure.

On univariate analysis, patients with past medical history of hydronephrosis had a 5-fold higher likelihood of developing a recurrence of hydronephrosis (45.9% [50/109] vs 8.4% [56/669], $P < 0.001$). Similarly, they also had a 9.5-fold higher likelihood of requiring urologic interventions related to the hydronephrosis (12.8% [14/109] vs 1.4% [9/669], $P < 0.001$). Having diagnoses predisposing the patient for urinary obstruction (benign prostate hyperplasia, abdominal/pelvic cancer, neurogenic bladder, single

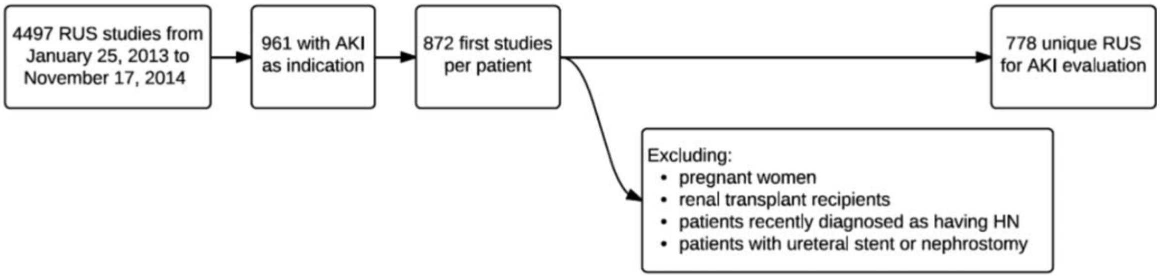

FIG. 1. Study cohort flow diagram.

TABLE 1. Patient Characteristics and Presence of Hydronephrosis on Renal Ultrasound

Patient Characteristic	With HN, n = 106	Without HN, n = 672	*P* Value
Demographics			
Age, y, mean ± SD	60.5 ± 17.1	64.1 ± 16.0	0.035*
Nonblack	97 (91.5)	573 (85.3)	0.084
Male	59 (55.7)	368 (54.8)	0.863
Past medical history			
Hydronephrosis	50 (47.2)	59 (8.8)	<0.001*
Recurrent urinary tract infections	22 (20.75)	101 (15.0)	0.133
Congestive heart failure	9 (5.5)	155 (23.1)	<0.001*
Prerenal status†	36 (34.0)	272 (40.5)	0.203
Exposure to nephrotoxic medication‡	20 (18.9)	260 (38.7)	<0.001*
Diagnosis consistent with obstruction□	59 (22.1)	208 (31.0)	<0.001*
Benign prostate hyperplasia	9 (8.5)	63 (9.4)	0.770
Abdominal or pelvic cancer	42 (39.6)	97 (14.4)	<0.001*
Neurogenic bladder	5 (4.7)	12 (1.8)	0.055
Single functional kidney	6 (18.8)	26 (81.3)	0.388
Pelvic surgery	14 (13.2)	61 (9.1)	0.181

NOTE: Data in parenthesis are percentages. **Abbreviations:** HN, hydronephrosis; SD, standard deviation. *Values are statistically significant. †Prerenal status: use of pressors or history of sepsis. ‡Nephrotoxic medications: aspirin (>81 mg/d), diuretic, angiotensin-converting enzyme inhibitor, or intravenous vancomycin. □Diagnosis consistent with possible obstruction: benign prostatic hyperplasia, abdominal or pelvic cancer, neurogenic bladder, single functional kidney, or previous pelvic surgery.

TABLE 2. Multivariable Model For Hydronephrosis Risk Stratification Among Patients With Acute Kidney Injury

Patient Characteristic	Adjusted Odds Ratio (95% Confidence Interval)	*P* Value
Race		
Nonblack (reference = black)	1.4 (0.7–3.1)	0.414
History of recurrent urinary tract infections		
Yes (reference = no)	0.75 (0.4–1.3)	0.346
Diagnosis consistent with possible obstruction*		
Yes (reference = no)	2.0 (1.2–3.1)	0.004†
History of HN		
Yes (reference = no)	7.4 (4.5–12.3)	<0.001†
History of CHF		
No (reference = yes)	2.7 (1.3–6.1)	0.009†
History of pre–renal AKI, use of pressors, or sepsis		
No (reference = 1)	1.0 (0.6–1.7)	0.846
Exposure to nephrotoxic medications prior to AKI‡		
No (reference = yes)	1.9 (1.1–3.3)	0.022†

NOTE: **Abbreviations:** AKI, acute kidney injury; CHF, congestive heart failure; HN, hydronephrosis.*Diagnosis consistent with possible obstruction: benign prostatic hyperplasia, abdominal or pelvic cancer, neurogenic bladder, single functional kidney, or previous pelvic surgery. †Values are statistically significant. ‡Nephrotoxic medications: aspirin (>81 mg/d), diuretic, angiotensin-converting enzyme inhibitor, or intravenous vancomycin.

functional kidney, and history of pelvic surgery) was correlated with the likelihood of both hydronephrosis and the need for urologic intervention. Of the patients with a diagnosis predisposing the patient for urinary obstructions, 22.1% (59/267) had hydronephrosis on imaging, whereas 9.2% (47/511) of patients without such a diagnosis had hydronephrosis ($P < 0.001$).

Conversely, having a recent exposure to nephrotoxic medications was negatively correlated with the likelihood of both hydronephrosis and the need for urologic intervention. Of the patients with recent exposure to nephrotoxic medications, 7.1% (20/280) had hydronephrosis on imaging, whereas the prevalence of hydronephrosis was 17.3% (86/498) in patients without such an exposure ($P < 0.001$) (Table 1).

Adjusted for other covariates, the multiple variable model showed that a diagnosis predisposing patients for obstruction (odds ratio [OR]: 2.0, $P = 0.004$), history of hydronephrosis (OR: 7.4, $P < 0.001$), absence of a history of congestive heart failure (OR: 2.7, $P = 0.009$), and lack of exposure to nephrotoxic medications (OR: 1.9, $P = 0.022$) were statistically significant predictors for hydronephrosis (Table 2).

After applying the Licurse renal ultrasonography risk stratification model, 176 (22.6%), 190 (24.4%), and 412 (53.0%) patients were classified as low risk, medium risk, and high risk for hydronephrosis, respectively. The incidence rates for hydronephrosis in the pretest probability risk groups were 4.0%, 6.8%, and 20.9% for low-, medium-, and high-risk patients, respectively ($P < 0.0001$). The rates for urologic interventions were 1.1%, 0.5%, and 4.9% in the risk groups from low to high ($P < 0.0001$) (Figure 2).

Overall, the Licurse model, using a cutoff between low-risk and medium/high-risk patients, had sensitivity of 91.3% (95% confidence interval [CI]: 73.2%-97.6%) for HNRI and 93.4% (95% CI: 87.0%-96.8%) for presence of HN. Specificity was low for both HNRI (23.0% [95% CI: 20.2%-26.2%]) and HN (25.1% [95% CI: 22.0%-28.6%]). The estimated potential reduction in renal ultrasound for hospitalized patients with AKI, defined as the rate of imaging performed in the low-risk group, was 22.6%. In the

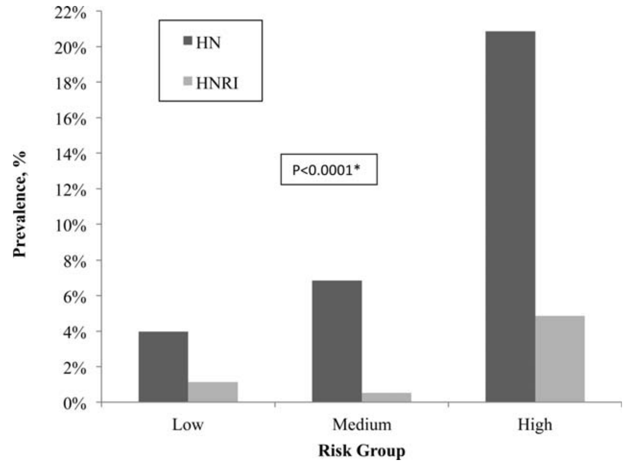

FIG. 2. Prevalence rates of hydronephrosis (HN) and hydronephrosis resulting in intervention (HNRI) across 3 risk stratification groups.

low-risk group, the number needed to screen to find 1 case of HN was 25, and to find 1 case of HNRI it was 88. The negative predictive value for hydronephrosis was 96.0% (95% CI: 92.0%-98.1%) and 98.9% for HNRI (95% CI: 96.0%-99.7%) (Table 3).

Supporting Table 1, in the online version of this article, shows a sensitivity analysis using different cut-off thresholds in the Licurse model for classifying low-risk patients. A lower threshold cutoff (ie, a cut-off of <1) significantly increases the sensitivity (98.1% [95% CI: 93.4%-99.5%] for HN; 100% [95% CI: 85.7%-100%]) for HNRI, but at the cost of a lower specificity (7.6% [95% CI: 5.8%-9.8%] for HN and 7.0% [95% CI: 5.4%-9.1%] for HNRI). The estimated potential reduction in renal ultrasound for hospitalized patients with AKI would be 6.0%, the number needed to screen to find 1 case of HN would be 26, and 1 case of HNRI would be infinity.

TABLE 3. Performance of Licurse Model on Patient Stratification in Validation Cohort

	Our External Validation Set		Licurse Internal Validation Set	
HN an Outcome	With HN	Without HN	With HN	Without HN
Low risk, no. of patients*	7	169	7	216
Medium/high risk, no. of patients	99	503	78	496
Test performance, % (95% CI)				
Sensitivity	93.4 (87.0–96.8)		91.8 (89.9–93.7)	
Specificity	25.1 (22.0–28.6)		30.3 (27.2–33.5)	
Negative predictive value	96.0 (92.0–98.1)		96.9 (95.7–98.1)	
HNRI an outcome				
Low risk, no. of patients	2	174	1	222
Medium/high risk, no. of patients	21	581	26	548
Test performance, % (95% CI)				
Sensitivity	91.3 (73.2–97.6)		96.3 (94.9–97.6)	
Specificity	23.0 (20.2–26.2)		28.8 (25.7–32.0)	
Negative predictive value	98.9 (96.0–99.7)		99.6 (99.1–100.0)	

NOTE: **Abbreviations:** CI = confidence interval; HN = hydronephrosis; NHRI = hydronephrosis requiring intervention. *Low-risk patients have <2 points on the Licurse model.

DISCUSSION

In this prospective observational study, we found that the Licurse risk stratification model, using a cutoff between low- risk and medium/high-risk patients, had 91.3% (95% CI: 73.2%-97.6%) sensitivity for predicting patients who would require urologic intervention and 93.4% (95% CI: 87.0%-96.8%) sensitivity for identifying patients with hydronephrosis. These findings were comparable to those found in the original validation cohort of the model, which showed sensitivity rates of 96.3% and 91.8%, respectively.[2] The negative predictive value for hydronephrosis and HNRI were sufficiently high, at 96.0% (95% CI: 92.0-98.1) and 98.9% (95% CI: 96.0-99.7), respectively.

Our results suggest that the Licurse model may be sufficient to rule out HN in the inpatient setting at our institution. The slight differences between the findings of our and the original studies may be due to differences in data extraction methodologies. In the original study, all data were retrospectively abstracted from medical records (discharge summaries and clinical notes) by 4 trained reviewers. However, such methodology is dependent on the quality of unstructured EHR data, which as noted in previous research, can be highly variable. Hogan and Wagner found that the correctness of EHR data can range from 44% to 100% and completeness from 1.1% to 100%, depending on the clinical concepts being studied.[4] Similarly, Thiru et al. found that the sensitivity of different types of EHR data ranged from 0.25 to 1.0.[5] Medical chart review can be labor intensive and time consuming. The lack of standardized methods for structured data capture has been a major limitation in decreasing research costs and speeding the rate of new medical discoveries through the secondary use of EHR data. By modifying our institutional clinical decision support (CDS) system to enable the necessary granular clinical data collection, we were able to obviate the need for resource intensive retrospective chart reviews. To our knowledge, this is the second example of a CDS tool specifically designed for capture of discrete data to validate a decision rule.[6] A similar process may also be useful to accelerate generation of new decision rules. With secondary use of EHR data becoming an increasingly important topic,[7] CDS may serve as an alternative method in the context of data reuse for clinical research. Based on a randomly selected chart review, it was noted that clinicians, overall, do try to communicate to the interpreting radiologists the clinical picture as accurately as they can, and rarely do providers drop their orders due to data entry.

Despite our data confirming Licurse's initial findings, it is important to note that as with any clinical prediction rules, there is a trade-off between cost savings and potential missed diagnoses. Even the most accepted clinical decision rules, such as the Well's criteria for pulmonary embolism and deep vein thrombosis, has their inherent "acceptable" rates of false negative. What is considered to be acceptable may differ among providers and patients. Thus, a shared decision-making model, in which the patient and provider actively engage in sharing of information regarding risks and benefits of both performing and bypassing the diagnostic testing, is preferred. For providers/patients who are more risk-adverse, one could consider using a more sensitive cutoff (for example, using the <1 threshold), essentially increasing the sensitivity from 91.3% to 100% for HNRI and from 93.4% to 98.1% for HN.

Although one would not want to miss a hydronephrosis in a patient, a too aggressive imaging strategy is not without economic and downstream risks. At an estimated cost of $200 per renal ultrasonography,[2] a 22.6% reduction would result in an annual savings of nearly $20,000 at our institution. The financial costs of forgoing ultrasound studies at the risk of missing 1 case of HN or 1 case of HNRI would be $5000 and $17,600, respectively.

Data-driven decision rules are becoming more commonly used in the current environment of increased emphasis on evidence-based medicine.[8–13] When applied appropriately, such prediction models can result in more efficient use of medical imaging while increasing value of care.[14,15] However, prior to implementation in clinical practice, these models need to be externally validated across multiple institutions and in various practice settings. This is the largest study of which we are aware to validate the utility of a prediction model for AKI in the inpatient setting. Although we did find slightly smaller differences in hydronephrosis in inpatients across the low, moderate, and high pretest probability groups, this may be explained by the differences in methodology.

Our study has several limitations. First, it was performed at a single academic medical center, a similar setting as that of the original work. Thus, the generalizability of our findings in other settings is unclear. Second, it is possible that our ordering providers did not thoroughly and accurately enter data into the structured CPOE form. However, we randomly selected a sample for chart review and found 90% concordance between data captured and those in the EHR. Due to selection of our cohort that included only patients with AKI who underwent RUS, it is possible that some patients who were not imaged or imaged with other cross-sectional modalities were excluded, resulting in differential test ordering bias. Finally, we did not include the potential benefits of RUS in affecting nonsurgical interventions of hydronephrosis (eg, Foley catheter insertion).

CONCLUSION

We found that the Licurse renal ultrasonography risk stratification model was sufficiently accurate in classifying patients at risk for ureteral obstruction among hospitalized patients with AKI.

Acknowledgements

The authors thank Laura E. Peterson, BSN, SM, for her assistance in editing this manuscript.

References

1. Gottlieb RH, Weinberg EP, Rubens DJ, Monk RD, Grossman EB. Renal sonography: can it be used more selectively in the setting of an elevated serum creatinine level? *Am J Kidney Dis.* 1997;29(3):362–367.
2. Licurse A. Renal ultrasonography in the evaluation of acute kidney injury: developing a Risk stratification framework. *Arch Intern Med.* 2010;170(21):1900.
3. Liu KD, Chertow GM. Curbing the use of ultrasonography in the diagnosis of acute kidney injury: Penny wise or pound foolish?: comment on "Renal ultrasonography in the evaluation of acute kidney injury." *Arch Intern Med.* 2010;170(21):1907–1908.
4. Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. *J Am Med Inform Assoc* 1997;4(5):342–355.
5. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. *BMJ.* 2003; 326(7398):1070.
6. Silveira PC, Ip IK, Goldhaber SZ, Piazza G, Benson CB, Khorasani R. Performance of Wells score for deep vein thrombosis in the inpatient setting. *JAMA Intern Med.* 2015;175(7):1112–1117.
7. Grande D, Mitra N, Shah A, Wan F, Asch DA. Public preferences about secondary uses of electronic health information. *JAMA Intern Med.* 2013;173(19):1798–1806.
8. Stiell IG, Wells GA, Vandemheen K, et al. The Canadian CT Head Rule for patients with minor head injury. *Lancet.* 2001;357(9266): 1391–1396.
9. Wells PS, Anderson DR, Bormanis J, et al. Value of assessment of pretest probability of deep-vein thrombosis in clinical management. *Lancet.* 1997;350(9094):1795–1798.
10. Dunning J, Daly JP, Lomas J-P, Lecky F, Batchelor J, Mackway-Jones K. Derivation of the children's head injury algorithm for the prediction of important clinical events decision rule for head injury in children. *Arch Dis Child.* 2006;91(11):885–891.
11. Perry JJ, Stiell IG, Sivilotti MLA, et al. Clinical decision rules to rule out subarachnoid hemorrhage for acute headache. *JAMA.* 2013; 310(12):1248–1255.
12. Wells PS, Anderson DR, Rodger M, et al. Excluding pulmonary embolism at the bedside without diagnostic imaging: management of patients with suspected pulmonary embolism presenting to the emergency department by using a simple clinical model and d-dimer. *Ann Intern Med.* 2001;135(2):98–107.
13. Stiell IG, Wells GA, Hoag RH, et al. Implementation of the Ottawa knee rule for the use of radiography in acute knee injuries. *JAMA.* 1997;278(23):2075–2079.
14. Ip IK, Schneider L, Seltzer S, et al. Impact of provider-led, technology-enabled radiology management program on imaging. *Am J Med.* 2013;126(8):687–692.
15. Raja AS, Ip IK, Prevedello LM, et al. Effect of computerized clinical decision support on the use and yield of CT pulmonary angiography in the emergency department. *Radiology.* 2012;262(2):468–474.