

**Critical Review Form
Diagnostic Test**

[Effect of provider experience on clinician-performed ultrasonography for hydronephrosis in patients with suspected renal colic. Ann Emerg Med. 2014 Sep;64\(3\):269-76.](#)

Objectives: "to determine the test characteristics of clinician-performed ultrasonography to detect hydronephrosis when performed by emergency clinicians with a wide range of experience, using noncontrast CT as a reference standard, and to determine whether there was an association between formal ultrasound (US) training and these test characteristics." (p. 270)

Methods: This prospective observation study was conducted at Yale-New Haven Hospital and Shoreline Medical Center (a freestanding hospital associated with Yale-New Haven Hospital) from July 19, 2010 to November 1, 2012. Consecutive emergency department (ED) patients aged 18 years and older presenting during predefined shifts for whom the provider caring for the patient intended to order a CT scan to evaluate for renal colic were eligible for inclusion. Exclusion criteria were performance of the CT with results already known, renal disease (chronic kidney disease, renal transplant, polycystic kidney disease, etc.), trauma, non-English speaking patient, incarceration, or inability to obtain consent. Subjects in whom the US was judged uninterpretable and those who did not undergo CT scan were excluded from analysis.

After enrollment patients underwent bedside ultrasonography, prior to CT scanning, by an emergency clinician, and the presence or absence of hydronephrosis was categorized as none observed, mild, moderate, or severe, as determined by the primary operator. The primary operator was either an emergency provider caring for the patient or an emergency physician on an US rotation. The level of training of the primary operator was categorized as one of the following:

- Attending physicians with fellowship training
- Attending physicians without fellowship training
- US experienced non-attending physician clinicians (required completion of at least 2 weeks of an US rotation)
- US inexperienced non-attending physician clinicians

All data was collected by "research assistants specifically trained in chart abstraction" (p. 271). All CT scans were interpreted by radiologists blinded to the results of the bedside ultrasonography. Hydronephrosis was considered to be present

if any degree of dilatation of the collecting system was noted in the dictated report; ureterolithiasis was considered present if a stone was detected anywhere between the renal pelvis and the ureterovesical junction.

There were 679 ED visits during the enrollment period by 672 unique individuals. In 9 cases the US was unable to be interpreted, leaving 670 total cases in the final analysis. The mean age was 46 years and 51.5% were female. On CT, 47% were found to have a ureteral stone seen and 47.4% had hydronephrosis. Interobserver agreement for presence of hydronephrosis on CT was excellent ($\kappa = 0.87$; 95% CI 0.73-1.00). The US was performed by a clinician caring for the patient in 542 (80.9%) of the 679 visits.

Guide		Comments
I.	Are the results valid?	
A.	Did clinicians face diagnostic uncertainty?	Yes. The presence of hydronephrosis was unclear at the time of bedside ultrasonography, which was performed prior to CT scan.
B.	Was there a blind comparison with an independent gold standard applied similarly to the treatment group and to the control group? (Confirmation Bias)	Yes. There were no treatment or control groups in the study, and included patients underwent both bedside US and CT scanning. The authors chose CT scan results as the gold standard, and specifically note that " the radiologist was blinded to the results of the bedside ultrasonography before interpreting the CT." (p. 271)
C.	Did the results of the test being evaluated influence the decision to perform the gold standard? (Ascertainment Bias)	No. To be eligible for enrollment, the treating physician must already have made the decision to perform a CT scan. Patients who did not undergo CT after enrollment were excluded from the analysis.
II.	What are the results?	
A.	What likelihood ratios were associated with the range of possible test results?	The overall test characteristics for any hydronephrosis on CT scan were LR+ of 2.72 (95% CI 2.25-3.27) and LR- of 0.37 (95% CI 0.31-0.44). For detection of moderate hydronephrosis on US vs. any hydronephrosis on CT, the LR+ was 5.76 (95% CI 3.61-9.19). The sensitivity, specificity, and positive likelihood ratios for different levels of training are detailed in Table 1 and Table 2.

		<p>Table 1. Test characteristics for moderate hydronephrosis on US vs. any hydronephrosis on CT based on level of experience</p> <table border="1"> <thead> <tr> <th></th> <th>Sensitivity (95% CI)</th> <th>Specificity (95% CI)</th> <th>LR+ (95% CI)</th> </tr> </thead> <tbody> <tr> <td>Attending with fellowship training</td> <td>38.2% (4.9-88.2)</td> <td>98.3 (82.8-99.9)</td> <td>22.52 (3.13-161.8)</td> </tr> <tr> <td>Attending without fellowship training</td> <td>23.1 (7.2-53.7)</td> <td>97.1 (89.7-99.2)</td> <td>8 (2.44-26.2)</td> </tr> <tr> <td>Experienced resident</td> <td>37.0 (19.6-58.7)</td> <td>90.8 (80.4-96.0)</td> <td>4.03 (2.12-7.65)</td> </tr> <tr> <td>Inexperienced clinician</td> <td>26.0 (10.4-51.6)</td> <td>93.8 (83.2-97.9)</td> <td>4.15 (1.64-10.51)</td> </tr> </tbody> </table> <p>Table 2. Test characteristics for any hydronephrosis on ultrasound vs. any hydronephrosis on CT based on level of experience</p> <table border="1"> <thead> <tr> <th></th> <th>Sensitivity (95% CI)</th> <th>Specificity (95% CI)</th> <th>LR+ (95% CI)</th> </tr> </thead> <tbody> <tr> <td>Attending with fellowship training</td> <td>92.7% (83.8-96.9)</td> <td>81.4% (63.8-91.5)</td> <td>4.97 (2.90-8.51)</td> </tr> <tr> <td>Attending without fellowship training</td> <td>61.5% (40.5-79.0)</td> <td>77.9 (59.9-89.2)</td> <td>2.78 (1.86-4.15)</td> </tr> <tr> <td>Experienced resident</td> <td>70.4 (59.3-79.5)</td> <td>70.6 (59.6-79.7)</td> <td>2.39 (1.74-3.28)</td> </tr> <tr> <td>Inexperienced clinician</td> <td>72.7 (54.4-85.7)</td> <td>65.0 (45.3-80.6)</td> <td>2.07 (1.49-2.88)</td> </tr> </tbody> </table>		Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	Attending with fellowship training	38.2% (4.9-88.2)	98.3 (82.8-99.9)	22.52 (3.13-161.8)	Attending without fellowship training	23.1 (7.2-53.7)	97.1 (89.7-99.2)	8 (2.44-26.2)	Experienced resident	37.0 (19.6-58.7)	90.8 (80.4-96.0)	4.03 (2.12-7.65)	Inexperienced clinician	26.0 (10.4-51.6)	93.8 (83.2-97.9)	4.15 (1.64-10.51)		Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	Attending with fellowship training	92.7% (83.8-96.9)	81.4% (63.8-91.5)	4.97 (2.90-8.51)	Attending without fellowship training	61.5% (40.5-79.0)	77.9 (59.9-89.2)	2.78 (1.86-4.15)	Experienced resident	70.4 (59.3-79.5)	70.6 (59.6-79.7)	2.39 (1.74-3.28)	Inexperienced clinician	72.7 (54.4-85.7)	65.0 (45.3-80.6)	2.07 (1.49-2.88)
	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)																																							
Attending with fellowship training	38.2% (4.9-88.2)	98.3 (82.8-99.9)	22.52 (3.13-161.8)																																							
Attending without fellowship training	23.1 (7.2-53.7)	97.1 (89.7-99.2)	8 (2.44-26.2)																																							
Experienced resident	37.0 (19.6-58.7)	90.8 (80.4-96.0)	4.03 (2.12-7.65)																																							
Inexperienced clinician	26.0 (10.4-51.6)	93.8 (83.2-97.9)	4.15 (1.64-10.51)																																							
	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)																																							
Attending with fellowship training	92.7% (83.8-96.9)	81.4% (63.8-91.5)	4.97 (2.90-8.51)																																							
Attending without fellowship training	61.5% (40.5-79.0)	77.9 (59.9-89.2)	2.78 (1.86-4.15)																																							
Experienced resident	70.4 (59.3-79.5)	70.6 (59.6-79.7)	2.39 (1.74-3.28)																																							
Inexperienced clinician	72.7 (54.4-85.7)	65.0 (45.3-80.6)	2.07 (1.49-2.88)																																							
III.	How can I apply the results to patient care?																																									
A.	Will the reproducibility of the test result and its interpretation be satisfactory in my clinical setting?	No. Based on the results of this study, the accuracy of point-of-care US for the detection of any hydronephrosis is poor in the hands of all clinicians, regardless of level of experience. While it could be argued that the presence of mild hydronephrosis in young patients with a ureteral stone is of limited clinical importance, the ability to verify the presumed diagnosis of ureteral stone is important. The positive LR _s for moderate hydronephrosis in the hands of attending physicians (both with and without fellowship training) are acceptable. The negative LR _s (while not provided in the study) are unfortunately quite poor (0.63 and 0.79 for fellowship and non-fellowship trained attending physicians, respectively).																																								
B.	Are the results applicable to the patients in my practice?	Yes. The study was conducted at a large academic ED associated with a residency training program and an associated freestanding ED. This is quite similar to our institution. Additionally, the patients in our practice would likely be similar to patients in the study.																																								
C.	Will the results change	No. I currently rarely perform bedside US in patients with a																																								

	my management strategy?	presumed diagnosis of ureteral colic. In rare situations where the diagnosis is either known or very likely, it would seem reasonable for an attending physician to perform a bedside US to evaluate for the presence of moderate hydronephrosis, with the understanding that the absence of moderate hydronephrosis on US does not rule out its presence.
D.	Will patients be better off as a result of the test?	Uncertain. This study did not evaluate outcomes. The diagnostic evaluation of ureteral colic has three goals: 1) confirmation of the presence of an obstructing ureteral stone, 2) evaluation for the presence and degree of hydronephrosis, and 3) exclusion of other potentially serious alternative causes of patients' symptoms. Given the relatively poor diagnostic characteristics of POC US for the diagnosis of any hydronephrosis, it seems that confirmation of ureteral colic as the cause of symptoms requires the presence of moderate hydronephrosis and an US performed by an attending physician. In all other cases, further work-up must be considered based on a) the level of suspicion that the patient has ureteral colic, and b) the risk of more serious alternative diagnoses.

Limitations:

- 1. The study was conducted at an academic center with strong ultrasound training. The results may not be applicable to community settings, or physicians without significant ultrasound training ([external validity](#)).**
- 2. Negative likelihood ratios were not provided for many of the analyses. The ability to reliably exclude hydronephrosis or ureteral colic from the diagnosis is as important as the ability to confirm the diagnosis.**
- 3. The authors calculated diagnostic test characteristics comparing moderate hydronephrosis found on US to any hydronephrosis seen on CT. This resulted in a falsely depressed false positive rate (and hence falsely increased specificity) at the expense of a falsely increased false negative rate (and hence falsely decreased sensitivity).**

Bottom Line:

This study suggests that bedside, clinician-performed ultrasound has, at best, moderate diagnostic accuracy. The authors report a high level of accuracy when performed by a fellowship trained attending physician, however the comparison of "moderate hydronephrosis" on US to "any hydronephrosis" on CT results in a falsely elevated specificity, albeit at the expense of sensitivity. It remains unclear how to clinically interpret the results of this study.