Prospective Evaluation of a Clinical Pathway for Suspected Appendicitis

WHAT’S KNOWN ON THIS SUBJECT: Although appendicitis is the most common surgical cause of abdominal pain in pediatrics, its diagnosis remains elusive. When evaluated independently, clinical scoring systems and ultrasonography have been shown to have low to moderate sensitivity in the diagnosis of appendicitis.

WHAT THIS STUDY ADDS: Our study evaluated the accuracy of a clinical practice guideline combining the Samuel’s pediatric appendicitis score and selective ultrasonography as the primary imaging modality for children with suspected appendicitis. Our clinical pathway demonstrated high sensitivity and specificity.

abstract

OBJECTIVE: To evaluate the diagnostic accuracy of a clinical pathway for suspected appendicitis combining the Samuel’s pediatric appendicitis score (PAS) and selective use of ultrasonography (US) as the primary imaging modality.

METHODS: Prospective, observational cohort study conducted at an urban, academic pediatric emergency department. After initial evaluation, patients were determined to be at low (PAS 1–3), intermediate (PAS 4–7), or high (PAS 8–10) risk for appendicitis. Low-risk patients were discharged with telephone follow-up. High-risk patients received immediate surgical consultation. Patients at intermediate risk for appendicitis underwent US.

RESULTS: Of the 196 patients enrolled, 65 (33.2%) had appendicitis. An initial PAS of 1–3 was noted in 44 (22.4%), 4–7 in 119 (60.7%), and 8–10 in 33 (16.9%) patients. Ultrasonography was performed in 128 (65.3%) patients, and 48 (37.5%) were positive. An abdominal computed tomography scan was requested by the surgical consultants in 13 (6.6%) patients. The negative appendectomy rate was 3 of 68 (4.4%). Follow-up was established on 190 of 196 (96.9%) patients. Overall diagnostic accuracy of the pathway was 94% (95% confidence interval [CI] 91%–97%) with a sensitivity of 92.3% (95% CI 83.0%–97.5%), specificity of 94.7% (95% CI 89.3%–97.8%), likelihood ratio (+) 17.3 (95% CI 8.4–35.6) and likelihood ratio (−) 0.08 (95% CI 0.04–0.19).

CONCLUSIONS: Our protocol demonstrates high sensitivity and specificity for diagnosis of appendicitis in children. Institutions should consider investing in resources that increase the availability of expertise in pediatric US. Standardization of care may decrease radiation exposure associated with use of computed tomography scans. Pediatrics 2014;133:e88–e95
Appendicitis is the most common surgical cause of atraumatic abdominal pain among children presenting to the emergency department (ED). Diagnosis of appendicitis by clinical examination alone remains elusive, and rates of perforated appendicitis in the pediatric population are high because its presentation overlaps with many other childhood illnesses that cause abdominal pain.\(^1\)\(^,\)\(^4\)

Early diagnosis of appendicitis is important because of the increased morbidity, mortality, and costs associated with perforated appendicitis.\(^5\)\(^,\)\(^6\) Although there is no diagnostic gold standard for appendicitis, 2 grading scores, the Alvarado and Samuel’s pediatric appendicitis score (PAS), have been developed to aid accurate diagnosis of appendicitis.\(^1\)\(^,\)\(^7\)\^-\(^9\)

The PAS is a score that was first reported by Samuel in *Journal of Pediatric Surgery* in 2002. Samuel’s score and PAS are used interchangeably (Table 1). A score of 1 to 3 is considered negative for appendicitis, whereas scores from 8 to 10 are considered positive. In his derivation study, Samuel did not precisely define percentage of neutrophilia or degree of elevation of temperature as a component of the PAS. We elected to use a differential count of 75% neutrophils or higher and a temperature of ≥38°C as an objective cutoff point. This is similar to other studies that validated the PAS.\(^8\)\(^,\)\(^10\)\(^,\)\(^11\) Both of these scoring systems are composed of 8 components, with a total score of 10.

The Alvarado score was initially developed in 1986 for use in the adult population. It has been validated in a subsequent study that included pediatric patients\(^1\)\(^,\)\(^4\) (Table 2). Alvarado scores of 1 to 4 are negative for appendicitis, whereas scores from 9 to 10 are considered diagnostic of appendicitis. Similar to the PAS, it also has 8 components with differences in definition of fever and descriptors for peritoneal signs on clinical examination.

The PAS was first published and oriented exclusively to the pediatric population. It has since been used in other studies that also demonstrated the limitations of exclusively using the PAS to identify patients with acute appendicitis.\(^8\)\(^,\)\(^10\)\(^,\)\(^11\)

We are not aware of any previous prospective studies that have used a clinical score and ultrasonography (US) for risk stratification of patients with abdominal pain with suspicion for appendicitis presenting to the ED.

The goal of the current study is to evaluate the diagnostic accuracy of a clinical pathway for suspected appendicitis using Samuel’s PAS and US as the primary imaging modality. Our hypothesis is that the sensitivity and specificity of the PAS with selective use of US would be superior to PAS alone.

**METHODS**

This was a prospective, observational study conducted at our urban, tertiary level, free-standing, pediatric ED with an annual census of 84,000 patient visits. Appropriate institutional review board approval with waiver of informed consent for completion of data forms and medical record review was obtained before the study initiation. We enrolled a convenience sample of patients between the ages of 3 and 17 years, presenting with abdominal pain and suspicion of appendicitis based on initial evaluation by the ED physician. We excluded patients with known inflammatory bowel disease, sickle cell disease, chronic steroids, or chronic immunosuppression. We also excluded patients who had a computed tomography (CT) scan of the abdomen before arrival at our institution and those who received antibiotics before arrival.

Our department’s usual practice for children presenting with suspected appendicitis is to administer a bolus of 20 mL/kg of isotonic intravenous fluids in conjunction with basic laboratory testing and period of observation in the ED. This includes a complete blood count, urinalysis, and metabolic panel. Also, based on the treating physician’s clinical judgment, they may receive a chest radiograph and/or a plain abdominal radiograph to exclude alternative diagnoses. Although our clinical pathway did not require a specific duration of observation, typically this period entailed time until completion of initial bolus and receipt of results of laboratory tests (Fig 1).

The PAS was assigned by the treating physician in the ED when results of complete blood count were available.

Our clinical pathway involved risk stratification based on the PAS. Patients with PAS of 1 to 3 (low probability of appendicitis) were either discharged from the hospital and received a follow-up phone call within 24 hours or, if warranted, admitted to the general pediatrics service with an alternate

---

**TABLE 1** Pediatric Appendicitis Score

<table>
<thead>
<tr>
<th>Sign/Symptom Description</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cough/percussion/heel tapping tenderness at RLQ</td>
<td>2</td>
</tr>
<tr>
<td>Anorexia</td>
<td>1</td>
</tr>
<tr>
<td>Low-grade fever ≥38°C</td>
<td>1</td>
</tr>
<tr>
<td>Nausea/emesis</td>
<td>1</td>
</tr>
<tr>
<td>RLQ tenderness on light palpation</td>
<td>2</td>
</tr>
<tr>
<td>Leucocytosis (&gt;10,000/mm(^3))</td>
<td>1</td>
</tr>
<tr>
<td>Left shift (&gt;75% neutrophilia)</td>
<td>1</td>
</tr>
<tr>
<td>Migration of pain to RLQ</td>
<td>1</td>
</tr>
<tr>
<td>RLQ: right lower quadrant</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 2** Alvarado Score

<table>
<thead>
<tr>
<th>Sign/Symptom Description</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Migration of pain</td>
<td>1</td>
</tr>
<tr>
<td>Anorexia</td>
<td>1</td>
</tr>
<tr>
<td>Nausea/vomiting</td>
<td>1</td>
</tr>
<tr>
<td>Right lower quadrant tenderness</td>
<td>2</td>
</tr>
<tr>
<td>Rebound pain</td>
<td>1</td>
</tr>
<tr>
<td>Increase in temperature (&gt;37.5°C)</td>
<td>1</td>
</tr>
<tr>
<td>Leucocytosis (&gt;10,000/mL)</td>
<td>2</td>
</tr>
<tr>
<td>Polymorphonuclear neutrophilia (&gt;75%)</td>
<td>1</td>
</tr>
</tbody>
</table>
diagnosis and without surgical consultation.

Participants with a PAS of 4 to 7 (intermediate probability of appendicitis) had a focused right lower quadrant US performed after a period of observation and parenteral hydration in the ED. The duration of observation and decision to obtain the US was left to the discretion of the treating clinician. If the US was negative and there remained no continued suspicion for appendicitis, the patient was discharged from the ED with a phone follow-up or admitted to the general pediatric service with an alternate diagnosis and without surgical consultation. If US was positive or there remained continued suspicion of appendicitis, surgical consultation was sought. If warranted, a CT scan was performed only after pediatric surgical consultation.

The written US report from the radiologist on call was used to aid medical decision-making. For the purpose of our study, US results were dichotomized as either positive or negative. If the appendix was visualized and reported as abnormal or if the report was suggestive of appendicitis on the basis of secondary signs of inflammation in the right lower quadrant, the result was deemed positive. The US was considered negative if the appendix was visualized and normal or if the appendix was not visualized and there were no secondary findings to suggest appendicitis.

The criteria for a positive or negative US were set a priori. Sonographic results were classified in a binary manner on the basis of evidence from the recent pediatric radiology literature examining methods to improve diagnostic performance of this modality in clinical practice. Specific radiologic criteria such as evidence of secondary signs of appendicitis, size of appendix, and presence of appendicolith were left to the discretion of the radiologist performing the US in real time. We also deliberately included preliminary, not final, radiology reports into our study database, to reflect information available at the time of clinical decision-making.

For participants with a PAS score of 8 to 10 (high probability of appendicitis), pediatric surgery was consulted for further management.

Use of CT scans to assist in the diagnosis of appendicitis was not a specific component of our clinical pathway guideline. Therefore, they were obtained only if requested by the consulting pediatric surgeon. The most common indications for a CT were either to provide additional information when the diagnosis was unclear or to assess for an intra-abdominal abscess, which may alter management approach.

Before enrollment of participants, the clinical pathway was presented at our physician staff meeting and monthly thereafter for the duration of the study. A copy of the algorithm was posted in the work area, and an electronic copy was shared with all ED physicians. The PAS and its components were built into our electronic medical record. The physicians could select the subcomponents, and a cumulative score would auto-populate in the patient’s record.

All the participating clinicians were advised to notify the principle investigator (AS) via e-mail or text within 24 hours of enrolling a patient in the pathway. In addition, the information systems analyst assigned to the ED provided the investigator (AS) with a weekly list of all patients who had a PAS recorded in their charts. After deidentification, subject data were
entered into a secure electronic spreadsheet for analysis.

Final follow-up was obtained through 3 mechanisms: operative and pathologic finding of appendicitis after surgical procedure, medical record review of hospital stay of patients admitted to the hospital for observation, and telephone follow-up at 24 hours after discharge of patients discharged from the ED. The gold standard used to confirm the presence of appendicitis was a pathology report consistent with appendiceal inflammation. Perforation was based on gross operative finding of a hole in the appendix as determined by the operating surgeon.

We used the following definitions to assess the diagnostic accuracy of our clinical pathway: patients were considered “test-positive” that is, high suspicion of appendicitis, if they had a PAS > 7 or a PAS of 4 to 7 and a US that was positive for appendicitis. Patients were considered “test-negative” if they had a PAS < 4 or a PAS of 4 to 7 and a negative US. Patients with a score of 4 to 7 who did not receive US because they improved after hydration or were noted to have an alternative diagnosis were also considered test-negative.

Data Analysis

The diagnostic accuracy of the clinical pathway was assessed by calculating its sensitivity, specificity and positive likelihood ratios along with 95% confidence intervals (CIs) using standard formulae. To compare the accuracy of the clinical pathway guideline, which is a dichotomous variable, to the PAS, which is a continuous variable, we fitted the receiver operator characteristic curve using the PAS score alone to our sample to obtain an optimal cutoff point. Sensitivity and specificity along with 95% CIs were calculated for the PAS score based on this optimal cutoff point. We then compared the diagnostic accuracy of the clinical pathway with that of the PAS score alone in our sample. All analyses were carried out using the software packages SAS version 9.3 (SAS Institute Inc, Cary, NC).

RESULTS

Two hundred sixteen patients were recruited over an 11-month period (October 2011–August 2012). We excluded 20 patients from analysis, 2 for incorrect enrollment because they had previous antibiotic use and 18 for protocol deviation (ie, imaging studies were obtained with a score < 4 or > 7, or surgical consultation was sought before advanced imaging for patients with a score of 4–7).

An initial PAS of 1 to 3 was noted in 44 (22.4%), 4 to 7 in 119 (60.7%), and 8 to 10 in 33 (16.9%) subjects. Of the 65 patients diagnosed with appendicitis, 0.0% had a low risk score, 37 (56.9%, 95% CI 44.4–69.2%) had an intermediate score, and 28 (43.1%, 95% CI 30.9–56.0%) had a high score. Of the patients with a low risk score, 0 of 44 (0.0%) had appendicitis. Of the patients with an intermediate score, 37 of 119 (31.1%) had appendicitis. Of the patients with a high-risk score, 28 of 33 (84.8%) had appendicitis. Perforated appendicitis was noted in 18 of 65 (15.4%) patients (Fig 2).

Ultrasonography was performed in 128 (65.3%) patients, of which 48 (37.5%) were positive for appendicitis. An abdominal CT scan was requested by the surgical consultants in 13 (6.6%) patients. Of the 58 patients who underwent an operation, 29 (42.6%) did not receive imaging. Ninety-nine of 196 patients were admitted for observation, and telephone follow-up at 24 hours after discharge of patients discharged from the ED. The gold standard used to confirm the presence of appendicitis was a pathology report consistent with appendiceal inflammation. Perforation was based on gross operative finding of a hole in the appendix as determined by the operating surgeon. Of the patients with an intermediate score, 37 of 119 (31.1%) had appendicitis. Of the patients with a high-risk score, 28 of 33 (84.8%) had appendicitis. Perforated appendicitis was noted in 18 of 65 (15.4%) patients (Fig 2).

Three of the 68 patients who underwent appendectomy had a normal appendix (4.4% 95% CI 0.09–12.4%). Two of these patients were admitted with an intermediate probability PAS and negative US. The other patient had a low-probability PAS, was discharged, and called back for continued severe abdominal pain. She was eventually diagnosed with omental infarction. One participant with an intermediate probability PAS score and negative US was discharged from the ED after she was deemed to have clinically improved with intravenous hydration. At the follow-up phone call the next day, she was advised to return to the ED for reevaluation and was eventually diagnosed with a ruptured appendix.

The optimal cutoff point for the PAS alone in our study was ascertained to be 6. At this cut point, the PAS score showed modest performance characteristics, with a sensitivity of 81.5% (95% CI 70.0–90.1%) and a specificity of 71.0% (95% CI 62.4–78.6%). The receiver operating characteristic curve with PAS alone at an optimal cutoff score of 6 is shown in Fig 3. The area under the curve is 0.8610 (0.8108–0.9111). In contrast, our clinical pathway had a sensitivity of 92.3% (95% CI 83.0%–97.5%), specificity of 94.7% (95% CI 89.3%–97.8%), likelihood ratio (+) 17.3 (95% CI 8.4–35.6) and likelihood ratios (−) 0.08 (95% CI 0.04–0.19; Table 3).

Median time to surgical consultation was 209.5 (interquartile range [IQR] 163.5–310.5) minutes from arrival at triage and 127.5 (IQR 79.0–182.5) minutes from initial evaluation by an ED physician. Median ED length of stay was 374 minutes (IQR 290.0–475.5). The CT use rate was 6.6% (13 of 196).

A summary of patient characteristics of the patients enrolled and those who were excluded is shown in Table 4. Eight patients who were noncompliant were diagnosed with appendicitis.
**DISCUSSION**

We prospectively evaluated a collaborative clinical pathway guideline, combining the PAS with selective use of US as the primary diagnostic imaging modality for patients with suspected appendicitis. Our results demonstrate that the diagnostic accuracy of our clinical pathway to risk-stratify patients with suspected appendicitis was superior to using the PAS alone, with significantly improved sensitivity and specificity. The likelihood ratio for a test enables a clinician to update his or her estimate of the probability of disease. Using our clinical pathway guideline, the likelihood of a patient with appendicitis having a positive “test” is 17.3 times greater than for a child without appendicitis. Conversely, the negative likelihood ratio of 0.08 tells us how much less likely it is that a child with appendicitis will test negative compared with someone without appendicitis.

Several studies have prospectively evaluated the Samuels and Alvarado scores in pediatric patients. Neither score was sufficient as a stand-alone to establish diagnosis of appendicitis. This dilemma has led to the recent trend of relying on diagnostic imaging in the evaluation of suspected pediatric appendicitis. CT scans, the imaging modality of choice, have improved diagnosis of appendicitis. As a result, use of CT scans for diagnosis of pediatric appendicitis has increased. A 10-year review of the National Ambulatory Medical Care Survey data in patients aged <19 years presenting to a pediatric ED noted a rise in CT use from 0.9% in 1998 to 15.4% in 2008. Furthermore, data from pediatric surgical services at 2 centers suggest that initial evaluation for suspected appendicitis at a community hospital is associated with a higher preoperative use of CT scans compared with a children's hospital (50%–75.2% vs 26.3%). Recently, because of heightened concerns surrounding risks of radiation exposure in children, US has emerged as an increasingly popular first-line diagnostic imaging modality, particularly at tertiary-level pediatric facilities where pediatric ultrasonographers are readily available. However, visualization of the appendix by US can be variable, potentially leading to many inconclusive studies. Nevertheless, despite judicious use of diagnostic imaging and the development of protocols for diagnosis of appendicitis, negative appendectomy rates in children remain high, ranging from 4.4% to 13%.

Few studies have systematically examined the performance characteristics of using a clinical pathway combining an objective appendicitis grading score with selective diagnostic imaging for children with suspected appendicitis. Our study has shown that use of a clinical pathway that combines a clinical grading score and selective use of US can
improve accuracy of risk stratification of suspected appendicitis while in the ED. Furthermore, this was accomplished while limiting CT scan use to 6.6% of our patients and maintaining a low rate of missed appendicitis and negative appendectomies.

In a recent study across Canadian pediatric EDs assessing site variations in flow metrics for children with suspected appendicitis, the average ED length of stay was 438 minutes, with a range of 321 to 638 minutes between their lowest and highest sites. We were able to keep the ED length of stay for patients in our study within the published length-of-stay metrics, despite having to bring in US technologists from home for imaging requests that occurred after regular working hours. Of note, 43% of the patients in our study population arrived in triage after 5 PM.

There are several limitations to our study. Without a comparative control group, we cannot objectively assess the impact of our clinical pathway on CT use and length of stay in the ED for patients with suspected appendicitis. Because we did not track patients with suspected appendicitis who were not enrolled during this period, it is possible that some patients with suspected appendicitis were evaluated in our ED and not enrolled in our study. Recent data suggest that the diagnostic value of a clinical score or laboratory test such as a complete blood count may vary at different time points of right lower quadrant pain. We did not specifically evaluate the duration of abdominal pain relative to the timing of imaging or laboratory tests.

It could be argued that we were assessing the impact of a suggested evaluation based simply on the PAS and clinical judgment rather than studying the impact of a pathway. Although clinical judgment was ineluctably linked to 2 subcomponents of the PAS (namely, assessment of right lower quadrant tenderness and presence or absence of peritoneal signs) the role of clinical judgment was minimized by adopting an objective score to risk-stratify patients, along with strict criteria for advanced imaging and surgical consultation. This decreased practice variation in the workup of patients with suspected acute appendicitis.
Our results cannot be generalized to a nonacademic and/or general ED, which may lack the around-the-clock availability of pediatric ultrasonographers and surgeons. We were unable to contact 6 patients at follow-up after being discharged from the ED.

The strength of our study was that we used a strict criterion standard and staged imaging protocol for evaluation of patients with suspected appendicitis. By using an objective, validated clinical scoring system, we may decrease variability in patient assessment among differing clinicians.

**CONCLUSIONS**

Our study suggests that a clinical pathway combining PAS and US for use in children with suspected appendicitis presenting to our pediatric ED demonstrates higher sensitivity and specificity than using the PAS alone. Institutions should consider investing in resources to improve availability and expertise in pediatric abdominal US, such that accuracy of diagnosis of appendicitis and minimization of radiation exposure can both be maintained in the pediatric population.

**ACKNOWLEDGMENTS**

We thank Sandy Grimes, RN, for her assistance in working with our institutional review board and all the physicians in the ED, Division of Surgery, and Department of Radiology at Le Bonheur Children’s Hospital for their support, without which the study would not have been feasible.

**REFERENCES**


Prospective Evaluation of a Clinical Pathway for Suspected Appendicitis
Ashley Saucier, Eunice Y. Huang, Chetachi A. Emeremni and Jay Pershad

Pediatrics; originally published online December 30, 2013;
DOI: 10.1542/peds.2013-2208
Prospective Evaluation of a Clinical Pathway for Suspected Appendicitis
Ashley Saucier, Eunice Y. Huang, Chetachi A. Emeremni and Jay Pershad

Pediatrics; originally published online December 30, 2013;
DOI: 10.1542/peds.2013-2208

The online version of this article, along with updated information and services, is located on the World Wide Web at:
/content/early/2013/12/24/peds.2013-2208