Validation of the Pediatric Appendicitis Risk Calculator (pARC) in a Community Emergency Department Setting

Dale M. Cotton, MD*; David R. Vinson, MD; Gabriela Vazquez-Benitez, PhD; E. Margaret Warton, MPH; Mary E. Reed, DrPH; Uli K. Chettipally, MD, MPH; Mamata V. Kene, MD, MPH; James S. Lin, MD; Dustin G. Mark, MD; Dana R. Sax, MD, MPH; Ian D. McLachlan, MD, MPH; Adina S. Rauchwerger, MPH; Laura E. Simon, BA; Anupam B. Kharbanda, MD, MSc; Elyse O. Kharbanda, MD, MPH; Dustin W. Ballard, MD, MBE; for the Clinical Research on Emergency Services and Treatments (CREST) Network

*Corresponding Author. E-mail: dale.m.cotton@kp.org.

Study objective: The pediatric Appendicitis Risk Calculator (pARC) is a validated clinical tool for assessing a child's probability of appendicitis. Our objective was to assess the performance of the pARC in community emergency departments (EDs) and to compare its performance with that of the Pediatric Appendicitis Score (PAS).

Methods: We conducted a prospective validation study from October 1, 2016, to April 30, 2018, in 11 community EDs serving general populations. Patients aged 5 to 20.9 years and with a chief complaint of abdominal pain and less than or equal to 5 days of right-sided or diffuse abdominal pain were eligible for study enrollment. Our primary outcome was the presence or absence of appendicitis within 7 days of the index visit. We reported performance characteristics and secondary outcomes by pARC risk strata and compared the receiver operator characteristic (ROC) curves of the PAS and pARC.

Results: We enrolled 2,089 patients with a mean age of 12.4 years, 46% of whom were male patients. Appendicitis was confirmed in 353 patients (16.9%), of whom 55 (15.6%) had perforated appendixes. Fifty-four percent of patients had very low (<5%) or low (5% to 14%) predicted risk, 43% had intermediate risk (15% to 84%), and 4% had high risk (≥85%). In the very-low- and low-risk groups, 1.4% and 3.0% of patients had appendicitis, respectively. The area under the ROC curve was 0.89 (95% confidence interval 0.87 to 0.92) for the pARC compared with 0.80 (95% confidence interval 0.77 to 0.82) for the PAS.

Conclusion: The pARC accurately assessed appendicitis risk for children aged 5 years and older in community EDs and the pARC outperformed the PAS. [Ann Emerg Med. 2019;■:1-10.]

Please see page XX for the Editor's Capsule Summary of this article.

0196-0644/$-see front matter

https://doi.org/10.1016/j.annemergmed.2019.04.023

INTRODUCTION

Pathologic inflammation of the appendix—appendicitis—is the most common surgical emergency in children.[1] Although treatment of appendicitis is effective, making the diagnosis in children can be challenging.[2-5] This diagnostic dilemma fuels practice variability and potentially increases unnecessary imaging.[6-10] Of specific concern, increased use of computed tomography (CT) scans exposes children to ionizing radiation and the potential risk of future malignancies without adequate evidence of improved appendicitis outcomes.[11-15] High use of CT is of particular concern in the community setting, in which, in a 2008 to 2012 review of greater than 2,500 pediatric patients, more than half of patients who underwent an appendectomy had a CT scan; and those in general EDs were approximately 8 times more likely to undergo CT imaging than a child evaluated in a facility with specialized pediatric resources.[16] Ultrasonography use shows promise and may mitigate overuse of CT, but its performance is operator and facility dependent.[17] Reliably identifying a patient's risk of appendicitis could improve care and reduce unnecessary imaging, as well as associated costs and complications. To this end, appendicitis risk scores have been developed and studied.

The pediatric Appendicitis Risk Calculator (pARC) is one such tool.[18] Variables include sex, age, duration of pain, guarding, pain migration, maximal tenderness in the right lower quadrant, and absolute neutrophil count. The pARC incorporates subtle and previously observed interactions by age and sex in regard to appendicitis risk and uses absolute neutrophil count on a continuous scale. Thus, the pARC requires use of an online or electronic health record integrated

Editor's Capsule Summary

What is already known on this topic

Both the pediatric Appendicitis Risk Calculator (pARC) and the Pediatric Appendicitis Score (PAS) are validated clinical prediction rules for appendicitis, but neither has been rigorously evaluated in community emergency departments (EDs).

What question this study addressed

How does the pARC compare with the PAS for the diagnosis of appendicitis?

What this study adds to our knowledge

In this prospective validation study conducted at 11 community EDs, the pARC was more accurate than the PAS for appendicitis (area under the curve 0.89, 95% CI 0.87 to 0.92 versus 0.80, 95% CI 0.77 to 0.82).

How this is relevant to clinical practice

The pARC outperformed the PAS and could be used to assist clinical decisionmaking for children with suspected appendicitis who presented to a general ED, although neither was tested against unstructured clinical gestalt.

calculator, differing from integer-tally scores such as the Alvarado or the Pediatric Appendicitis Score (PAS).[19,20] In its previous validation in an academic children's hospital cohort with a 35% baseline risk for appendicitis, the pARC was able to accurately sort nearly half of patients into a category of either low (<15%) or high (≥85%) probability of appendicitis.[18] This ability to discriminate high- and low-risk strata is an improvement over other scores because accurate classification at either end of the risk spectrum may obviate the need for imaging. Yet, like many pediatric clinical decision tools, those for pediatric appendicitis have rarely been studied in the community setting in which they are most likely to be used.[21,22] Validation in the setting in which a tool is used is a core tenet of a robust decision tool.[22] The aims of the current study were to evaluate the performance of pARC when used in a community emergency department (ED) setting and to compare its performance with that of the PAS.

MATERIALS AND METHODS

Study Design and Setting

We conducted this multicenter, prospective, observational cohort validation study of the pARC from October 1, 2016, to April 30, 2018, in 11 community EDs serving general populations in Kaiser Permanente Northern California. These study centers are a subset of a larger 17-center cluster-randomized trial of electronic clinical decision support to aid in the diagnostic evaluation of children with acute abdominal pain. Kaiser Permanente Northern California is an integrated health care delivery system that provides comprehensive medical care for greater than 4 million members, who are representative of the ethnic and socioeconomic diversity of the surrounding population.[23] It is a learning health care system with an applied research agenda, and is supported by a comprehensive integrated electronic health record (Epic, Verona, WI), which includes inpatient, outpatient, emergency, pharmacy, laboratory, and imaging data.[24,25] The Kaiser Permanente Northern California Institutional Review Board approved the study with a waiver of informed consent.

The 2017 cumulative annual census of the 11 participating EDs was 793,000, with greater than 136,000 patients (17.2%) in our study aged 5 to 20.9 years (for facility characteristics, see Table E1, available online at http://www.annemergmed.com). None of these facilities are university based, but 5 have academic emergency medicine affiliations and 4 are referral centers for pediatric appendectomies. Referral centers staffed pediatric surgeons and nonreferral centers varied in regard to the age of patients transferred to Kaiser Permanente Northern California referral centers for appendectomy. All EDs were staffed by general emergency medicine residency–trained (board-certified or board-eligible) physicians. EDs had 24-hour access to CT imaging and variable but daily access to ultrasonography.

Selection of Participants

We included ED patients aged 5 to 20.9 years with a chief complaint of recent-onset (≤ 120 hours) generalized or right-sided abdominal pain. Physicians were trained by a local study champion to enroll eligible patients into a Web-services-based clinical decision support system from within the ED Navigator menu of the electronic health record. This Kaiser Permanente Northern California clinical decision support system has been successfully used in other clinical applications.[26,27] In addition to clinical decision support system familiarity from previous studies and ongoing education by a study champion, we began sending electronic text alerts to physicians' smartphones 3 months into the study period to notify them that a patient assigned to them in the electronic health record might be study eligible.[28]

As in the derivation and validation studies, patients were ineligible for enrollment if they had any of the following: abdominal trauma within 7 days, current pregnancy,

history of abdominal surgery including appendectomy, inflammatory bowel disease, chronic pancreatitis, sickle cell anemia, cystic fibrosis, and other conditions that might affect the ability to obtain an accurate history or physical examination (for exclusionary conditions and associated International Classification of Diseases codes, see Table E2, available online at http://www.annemergmed.com). In cases of multiple enrollments caused by more than one visit, only the first enrollment was included. The clinical decision support system was open to advanced practitioners and trainees, but we limited data inclusion to those entered by attending physicians.

Data Collection and Processing

After opening the clinical decision support system, physicians were presented with prepopulated patient-specific eligibility criteria for editing and confirmation (screen shots of the clinical decision support system are shown in Figure E1, available online at http://www.annemergmed.com). If the patient met study criteria, the physician advanced the clinical decision support system to the data collection screen to input 7 variables from the history and physical examination. For this analysis, pARC scores were calculated post hoc for patients for whom a white blood cell (WBC) count was obtained in the ED and when pARC clinical decision support system clinical data were otherwise complete. At this point as part of the parent study, physicians at certain intervention sites were presented the pARC score, and others at control sites were not.

Outcome Measures

Our primary outcome measure was the presence or absence of acute appendicitis within 7 days of the index visit. Acute appendicitis was identified by screening for 7 days from the index visit for an ED or hospital diagnosis of appendicitis, current procedural terminology code for appendectomy, or both. Our outcomes verification process reflected principles established for robustness in research involving chart review.[29] Outcome reviewers were blinded to the pARC score and the PAS. If operative and pathology reports were available, outcome verification was performed by text string search algorithms and manual chart review. If these 2 methods were discordant or ambiguous, then a second reviewer assessed the data. Discordant manual assessments were adjudicated by a third reviewer. If manual review confirmed that no operative or pathology reports were available, then the chart, imaging, and discharge medication list were reviewed to identify the impression of acute appendicitis with a nonoperative treatment plan. An imaging or ED diagnosis alone of appendicitis was insufficient to determine the outcome as appendicitis.

Our secondary outcomes were appendiceal perforation, negative appendectomy, and missed appendicitis within 7 days of the index visit. Appendiceal perforation for patients undergoing appendectomy was determined by the surgeon's intraoperative note for key words such as *presence of abscess*, *peritonitis*, *complex appendicitis*, or *purulent material*. Enrolled patients who were discharged after their index ED visit and subsequently met our study definition of acute appendicitis within 7 days were considered to be a missed appendicitis case. We searched our claims database for potential missed appendicitis-related health care visits outside of Kaiser Permanente Northern California whose index visit may have been within the delivery system. Cases of missed appendicitis were adjudicated by 4 study investigators (L.E.S., A.B.K., E.O.K., and D.W.B.). Negative appendectomy results were defined as appendectomies in which the pathology revealed no evidence of appendiceal inflammation.

We tracked appendectomy procedures and principal diagnosis of appendicitis for patients who were eligible to be enrolled but were not. We report the size of this cohort and the presumed appendicitis rate.

For each patient, we calculated the pARC score and the PAS (for pARC and PAS variables, see Table E3, available online at http://www.annemergmed.com). The pARC variables include sex, age, duration of pain, guarding, pain migration, maximal tenderness in the right lower quadrant, and absolute neutrophil count.[18] When absolute neutrophil count was unavailable (5% of cases), it was estimated from the WBC count (see Table E4, available online at http://www.annemergmed.com).[18,28] The PAS variables include cough, percussion, and hopping tenderness in the right lower quadrant; anorexia; pyrexia; nausea or emesis; tenderness over the right iliac fossa; leukocytosis; neutrophilia; and migration of pain to the right lower quadrant.[30] The pARC differs from the PAS in that it provides the risk of appendicitis on a continuous scale. The pARC was developed in a sample of children with suspected appendicitis defined as undergoing laboratory testing, diagnostic imaging, or a surgical consultation for appendicitis in patients aged 5 to 18 years, with a 40% rate of appendicitis.[18] The pARC equation can be found in Table E4 (available online at http://www.annemergmed.com).

We sorted pARC scores into 1 of 7 clinically actionable risk strata: less than 5%, 5% to 14%, 15% to 24%, 25% to 49%, 50% to 74%, 75% to 84%, and greater than or equal to 85%. Qualitatively, we described the less than 5% group as very low risk, 5% to 14% as low risk, 15% to 84% as intermediate risk, and greater than or equal to 85% as high

risk. These strata were chosen by multidisciplinary study-team consensus as having distinct diagnostic or management approaches.

We report discriminatory performance features (sensitivity, specificity, positive predictive value, negative predictive value, and positive and negative likelihood ratios) and secondary outcomes (missed appendicitis, perforation, and negative appendectomy) by risk strata. Overall discriminatory performance was evaluated and compared for the pARC and the PAS by generating the receiver operating characteristic curve and the area under the curve (AUC) statistic. We also report the AUC statistic range by facility. To best represent the discriminatory performance of the pARC, we initially excluded patients transferred outside of Kaiser Permanente Northern California. In sensitivity analysis, we included these transfer patients with a presumed diagnosis of appendicitis. As part of validating the pARC in this newly studied population, we estimated the calibration intercept and slope. This was achieved by regressing the logit of pARC values to the observed appendicitis outcome with logistic regression,[31] plotting the observed and predicted appendicitis risk from pARC and calibrating pARC according to a decile partition of the distribution of pARC (calibration plot), and computing the Hosmer-Lemeshow goodness-of-fit test statistic. In addition, we estimated the latter test statistic for the PAS.[32,33] The data analysis was generated with SAS/STAT software (version 9.4; SAS Institute, Inc., Cary, NC).

We manually reviewed all cases of all patients with a very low pARC score (<5%) who received a diagnosis of appendicitis. We present the collected variables for each of these cases, as well as the pARC score and the presence of perforation or missed appendicitis.

RESULTS

Characteristics of Study Subjects

During the 18-month study period, our study included 2,089 patients from 11 community facilities with a median of 151 patients (interquartile range [IQR] 107 to 283 patients) per facility. Four hundred fifteen providers enrolled patients, with a median of 4 enrollees per provider (IQR 2 to 7 enrollees). The median patient age was 12 years (IQR 9 to 16 years). Clinical characteristics and study flow are provided in Table 1 and Figure 1, respectively. Of enrollees, 46% were male patients and 56% presented with less than 24 hours of pain. The most commonly reported symptoms were nausea or vomiting (69%). The median WBC count was 9.9x10^3/μL (IQR 7.5x10^3/μL to 13.5x10^3/μL). Appendicitis was confirmed in 353 cases (16.9%).

Table 1. Study population characteristics.

Characteristic	Cohort, No. (%), N = 2,089
Median age (IQR), y	12 (9–16)
Age, No. (%), y	
Male patients	963 (46)
5–7.9	221 (11)
8–13.9	446 (21)
14–20.9	296 (14)
Female patients	1,026 (54)
5–7.9	173 (8)
8–11.9	300 (14)
12–20.9	653 (31)
Reported clinical presentation, No. (%)	
Duration of abdominal pain, h	
<24	1,179 (56)
24–47	307 (15)
48–120	603 (29)
Nausea or vomiting	1,436 (69)
Pain with walking or hopping	1,004 (48)
Migration of pain to RLQ	660 (32)
Findings on examination, No. (%)	
Maximal tenderness in RLQ	1,009 (48)
Abdominal guarding	524 (25)
Median laboratory results (IQR)	
WBC, ×10^3/μL	9.9 (7.5–13.5)
ANC, ×10^3/μL	6.8 (4.43–10.53)
Appendicitis confirmed, No. (%)	353 (17)

RLQ, Right lower quadrant; *ANC*, absolute neutrophil count.

Performance characteristics for the pARC are shown in Table 2 and Table E5 (available online at http://www.annemergmed.com). Fifty-four percent of patients had a low- or very-low-risk score and 4% had a high-risk score. The observed rate of appendicitis was 1.4% and 3.0% within the very-low- and low-risk strata, respectively, and 84.8% in the high-risk strata. The 2 lowest pARC strata had very high sensitivity, 100.0% and 97.5% for pARC score less than 5% and 5% to 14%, respectively. The 2 highest pARC strata had very high specificity, 97.8% and 99.3% for pARC score 65% to 84% and greater than or equal to 85%, respectively.

The overall pARC discriminatory performance was high, with an AUC of 0.89 (95% confidence interval [CI] 0.87 to 0.92), which was better than the PAS AUC of 0.80 (95% CI 0.77 to 0.82). Compared with the PAS at any specificity, pARC sensitivity was higher (Figure 2). We also calculated discriminatory performance for both scores, treating all patients transferred out of system (20 patients)

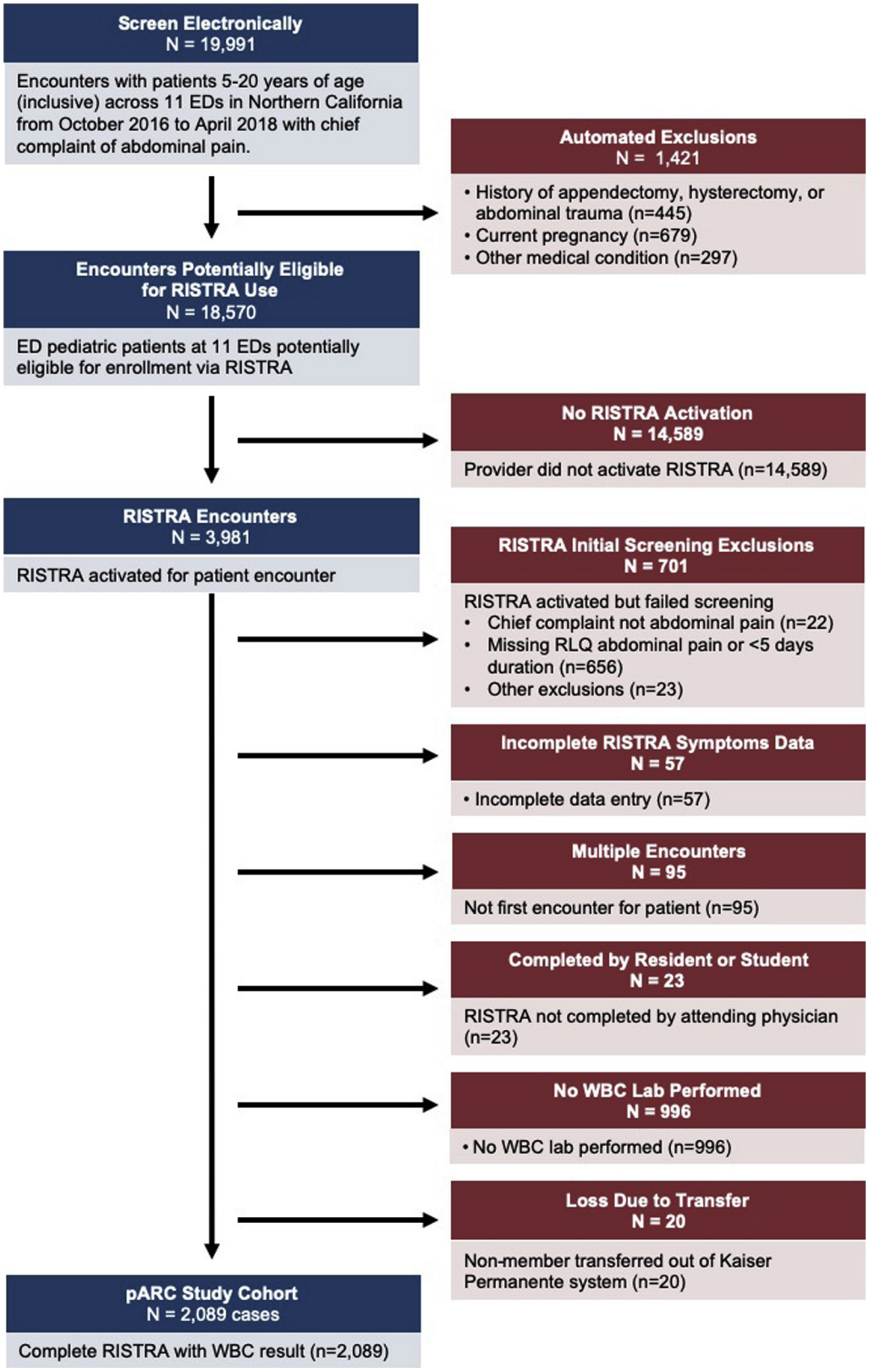

Figure 1. Flow diagram. *RISTRA*, Risk stratification tool.

Table 2. pARC performance and safety outcomes by predicted probability of appendicitis.

pARC Score,* %		Number of Enrollees (% of Total)	Appendicitis Cases	Appendicitis Prevalence, % (95% CI)	Sensitivity,† % (95% CI)	Specificity,† % (95% CI)	Missed Appendicitis,‡ No. (%)	Perforation, No. (%)	Negative Appendectomy, No. (%)
<5	Very low	661 (32)	9	1.4 (0.5–2.3)	100.0	0	2 (0.3)	0	4 (30.8)
5–14	Low	462 (22)	14	3.0 (1.4–4.6)	97.5 (95.9–99.1)	37.6 (35.3–39.9)	0	0	4 (22.2)
15–24	Intermediate	247 (12)	19	7.7 (4.4–11.0)	93.5 (90.9–96.1)	63.4 (61.1–65.7)	1 (0.4)	1 (5.3)	3 (13.6)
25–49		335 (16)	80	23.9 (19.3–28.5)	88.1 (84.7–91.5)	76.5 (74.5–78.5)	2 (0.6)	10 (12.5)	8 (9.1)
50–74		230 (11)	115	50.0 (43.5–56.5)	65.4 (60.4–70.4)	91.2 (89.9–92.5)	4 (1.7)	23 (20.0)	3 (2.5)
75–84		75 (4)	49	65.3 (54.5–76.1)	32.9 (28.0–37.8)	97.8 (97.1–98.5)	0	9 (18.4)	0
≥85	High	79 (4)	67	84.8 (76.8–92.8)	19.0 (14.9–23.1)	99.3 (98.9–99.7)	1 (1.3)	12 (17.9)	1 (1.5)
Total		2,089	353				10	55	23

*Predicted probability of appendicitis.
†The test result was positive if the score was greater than or equal to certain cut points (≥0, ≥5, ≥15, ≥25, ≥50, ≥75, and ≥85).
‡Missed appendicitis percentage as missed appendicitis cases per total number of patients in that risk strata.

as presumed appendicitis cases instead of excluding them. No change in discriminatory performance was noted. Across facilities, the pARC AUC ranged from 0.85 to 0.94. The pARC demonstrated adequate calibration, as shown in Figure 3. Calibration intercept was –0.615, slope was 1.10, and Hosmer-Lemeshow goodness-of-fit test result was 11.81 (8 *df*; *P*=.16). The PAS Hosmer-Lemeshow goodness-of-fit test result was 2.4 (5 *df*; *P*=.78).

The negative appendectomy rate was 6.5%. It was highest for the very low and low pARC scores (30.8% and 22.2%, respectively) and 0.9% in the 2 highest-risk strata combined (Table 2). The overall perforation rate was 15.6% and there were no perforations in the 2 lowest-risk strata. There were 9 cases of missed appendicitis, 2 of which were in the very-low-risk pARC strata.

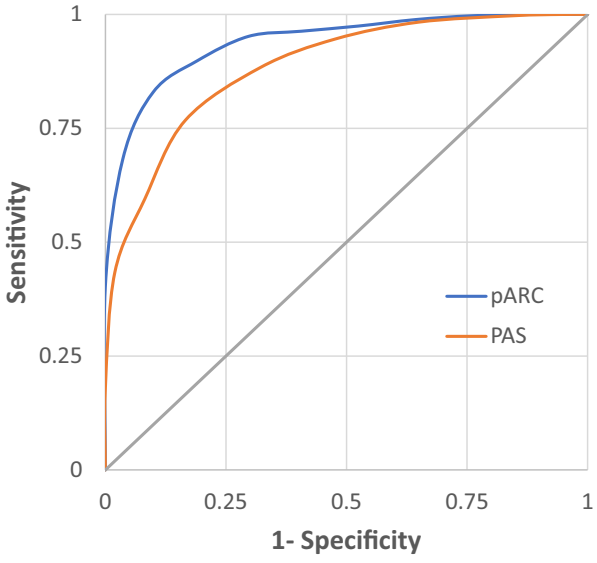

Figure 2. Receiver operator curve for pARC and PAS. pARC AUC=0.89 (95% CI 0.87 to 0.92); PAS AUC=0.80 (95% CI 0.77 to 0.82).

During our study period, there were 14,589 patients aged 5 to 20.9 years with a chief complaint of abdominal pain who received care in our 11 participating EDs but were not enrolled in our study. Of these patients, 126 (0.86%) had appendicitis. Demographic data comparing the enrolled and nonenrolled cohorts are available in Table E6 (available online at http://www.annemergmed.com).

Nine patients (1.4%) with pARC scores less than 5% had appendicitis (Table 3). Six were female patients, and 6 patients reported less than 24 hours of abdominal pain at presentation. Several patients had postenrollment

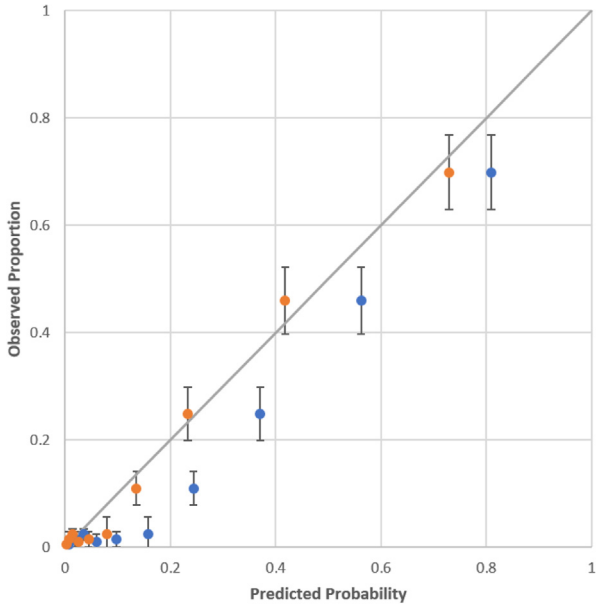

Figure 3. Calibration plot for pARC. Blue plot is pARC. Orange plot is pARC after calibration. Calibration intercept was –0.615, slope was 1.10, and Hosmer-Lemeshow goodness-of-fit test result was 11.81 (8 df; *P*=.16).

Table 3. Characteristics of 9 patients with very low pARC scores who received a diagnosis of appendicitis.

Case No.	Age, Years	Sex	Duration of Pain, Hours	Walking	Migration	Maximal Tenderness in the RLQ	Guarding	WBC Count (ANC)	Perforation	Missed Appendicitis	pARC Score,* %
1	13	F	<24	No	Yes	Yes	No	7.1 (3.4)	No	Yes	2.1
2	19	M	<24	Yes	No	No	No	5.0 (3.0)	No	No	3.6
3	16	F	48–120	Yes	Yes	Yes	No	5.6 (2.1)	No	No	3.4
4	7	M	48–120	No	No	No	No	6.5 (5.7)	No	No	2.3
5	17	F	<24	No	Yes	No	No	6.9 (3.9)	No	No	0.9
6	19	F	<24	Yes	No	No	No	8.8 (6.3)	No	No	4.0
7	12	M	<24	No	No	No	No	7.0 (4.3)	No	Yes	2.2
8	20	F	48–120	No	Yes	Yes	Yes	7.1 (3.3)	No	No	4.3
9	8	F	<24	No	No	No	No	9.4 (7.2)	No	No	3.7

F, Female; *M*, male.
*Predicted probability of appendicitis.

documentation that suggested a progressing examination, but only one had clear documentation that would have moved the patient to an intermediate-risk pARC score, if recalculated. All 9 patients had a normal WBC count (3.7 to $11.1\times10^3/\mu L$) and absolute neutrophil count (1.8 to $7.9\times10^3/\mu L$) for our system reference. Two of these cases were missed appendicitis; both patients had less than 24 hours of symptoms. There were no cases of appendiceal perforation among patients with a very low pARC score.

LIMITATIONS

There are several limitations specific to our study. Although we have established that the pARC is accurate and safe for sorting patients into risk strata for appendicitis, we have not yet shown that it can be used to improve clinical care, such as decreasing the rate of CT scans. Work to demonstrate improved imaging use is ongoing by our study team. Additionally, in contrast to previously studied integer-tally risk scores such as the PAS, the pARC requires the use of a calculator. Although these calculations could be achieved by integration with the electronic health record or Web-based tools, it may prove a barrier in some clinical environments.[34] If implementation hurdles can be overcome, there is evidence of substantial benefit supporting the efficacy of a clinical decision support system.[27,35-38]

Although the size of our cohort and the setting from which it was derived represent a substantial sample, our results may not be applicable to unique practice environments. As in the original pARC validation study, we did not include patients younger than 5 years because of the inconsistent and protean manifestations of appendicitis in the very young.[18,39,40] Despite these limitations, our pARC validation would apply to the majority of pediatric patients in the United States who present to the ED with possible appendicitis.[23]

DISCUSSION

In this study of the pARC performance within an integrated community-based health care system, pARC was validated as a reliable tool for classifying patients with suspected appendicitis into clinically relevant risk strata. The risk score did so while outperforming the PAS in terms of discrimination and overall model performance. This validation study may provide clinicians working in community EDs serving a general population with confidence that the pARC can be used to reliably estimate appendicitis risk.

Multiple clinical decision tools to aid in the diagnosis of pediatric appendicitis have been described.[41,42] They vary widely in their congruence to published criteria for rigor in clinical decision rules.[22,43] Two of the more well-studied scores include the PAS and Alvarado scores.[8,44-49] Investigations of these tools have shown methodological or performance features that limit clinical application.[20,50-52] Specifically, these tools generally place the majority of patients into intermediate-risk categories that indicate imaging use; this is in contrast to the pARC, in which the majority of patients are sorted into low- and high-risk categories that could help mitigate imaging use. In addition, these decisions tools have rarely been studied in the setting in which the majority of children present: community EDs.[21]

Another common shortcoming of clinical decision tools is capturing only a subset of the intended study population. This is frequently observed in studies of pediatric appendicitis that enroll patients already admitted to the

hospital or after appendectomy has already been decided on, far past a point at which diagnostic decisions have been made by the initial provider.[49] Such selection biases may miss true appendicitis cases, affect the performance characteristics of the tool, limit its generalizability, and likely contribute meaningfully to the heterogeneity of reported rates of appendicitis.[53] Previous work has demonstrated that our triage identification of patients by a chief complaint of abdominal pain has a sensitivity of 97% for pediatric appendicitis cases presenting to the ED.[54] Many of our study-eligible patients were not enrolled in the clinical decision support system by the physician, possibly because of low suspicion for appendicitis, given that this eligible but nonenrolled cohort had a very low appendicitis rate of 0.86%. Because our triage identification was highly sensitive and our missed eligible appendicitis rate was very low, we believe we captured the majority of appendicitis cases presenting to the ED during the study period.

Our study adds to the existing literature on this topic in several ways. First, we studied a diverse group of pediatric patients, representative of the spectrum of pediatric patients presenting for acute care in the United States. Second, the discriminatory performance of the pARC was found to be higher than that of the previously studied PAS. Third, we described how the pARC provided more clinically actionable information than the PAS by stratifying the majority of patients into low- or high-risk strata. Fourth, the pARC provided these performance characteristics while maintaining patient safety, demonstrated by a negative appendectomy rate of 6.5% and perforation rate of 15.6%, which are comparable to those in published reviews of pediatric populations.[5,12,55,56]

There are exciting opportunities for further work to improve the pARC. One is to leverage the pARC's dependence on an electronic calculator to allow dynamic calibration of the score according to the appendicitis prevalence or practice setting in which a patient is being assessed for appendicitis. For example, when initiating the calculator, a user might select an academic children's hospital setting (as in the original validation study) or community setting (as in this study) to set the calibration.[31] Such dynamic calibration is not necessary for utility because we have shown the pARC's clinically useful performance without it, but it may offer a path to further improve the tool's risk assessment. Other avenues may include risk stratifying patients for antibiotic-only treatment of appendicitis or studying pARC performance in non-ED settings such as urgent care.[57]

In this external validation study of more than 2,000 pediatric patients presenting for care in 11 community EDs, we have shown that the pARC safely and accurately assesses appendicitis risk for children aged 5 years and older who may have appendicitis. Further study is needed to understand the effect of pARC on the clinical care of patients with suspected appendicitis.

Supervising editor: Lise E. Nigrovic, MD, MPH. Specific detailed information about possible conflict of interest for individual editors is available at https://www.annemergmed.com/editors.

Author affiliations: From the Permanente Medical Group, Oakland, CA (Cotton, Vinson, Chettipally, Kene, Lin, Mark, Sax, McLachlan, Ballard); Kaiser Permanente, South Sacramento Medical Center, Sacramento, CA (Cotton); Kaiser Permanente, Division of Research, Oakland, CA (Vinson, Warton, Mark, Rauchwerger, Simon, Ballard); Kaiser Permanente, Sacramento Medical Center, Sacramento, CA (Vinson); Health Partners Institute, Division of Research, Bloomington, MN (Vazquez-Benitez, E. O. Kharbanda); Kaiser Permanente, South San Francisco Medical Center, South San Francisco, CA (Chettipally); Kaiser Permanente, San Leandro Medical Center, San Leandro, CA (Kene); Kaiser Permanente, Santa Clara Medical Center, Santa Clara, CA (Lin); Kaiser Permanente, Oakland Medical Center, Oakland, CA (Mark, Sax); Kaiser Permanente, San Francisco Medical Center, San Francisco, CA (McLachlan); Children's Hospitals and Clinics of Minnesota, Department of Pediatric Emergency Medicine, Minneapolis, MN (A. B. Kharbanda); and Kaiser Permanente, San Rafael Medical Center, San Rafael, CA (Ballard).

Author contributions: DMC, DRV, MER, ABK, EOK, and DWB conceived the study and its design. DRV, DGM, ASR, ABK, EOK, DWB, and MER obtained research funding. DMC, LES, ABK, EOK, and DWB performed manual chart review. GV-B and EMW performed the programming and analysis. DMC, DWB, GVB, MER, EMW, ABK, and EOK interpeted the data. DWB oversaw the study as a whole. DMC drafted the article and all authors contributed substantively to its critical revision and its final approval. DMC takes responsibility for the paper as a whole.

All authors attest to meeting the four ICMJE.org authorship criteria: (1) Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND (2) Drafting the work or revising it critically for important intellectual content; AND (3) Final approval of the version to be published; AND (4) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding and support: By *Annals* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist. Supported by the National Institutes of Health (R01 HD079463).

Publication dates: Received for publication December 12, 2018. Revision received March 21, 2019. Accepted for publication April 19, 2019.

Presented at the American College of Emergency Physicians (ACEP) meeting, October 2018, San Diego, CA and Health Care

Systems Research Network annual meeting, April 2019, Portland, OR.

Trial registration number: NCT02633735

REFERENCES

1. Scholer SJ, Pituch K, Orr DP, et al. Clinical outcomes of children with acute abdominal pain. *Pediatrics*. 1996;98(4 pt 1):680-685.
2. Rothrock SG, Skeoch G, Rush JJ, et al. Clinical features of misdiagnosed appendicitis in children. *Ann Emerg Med*. 1991;20:45-50.
3. Rasmussen T, Fonnes S, Rosenberg J. Long-term complications of appendectomy: a systematic review. *Scand J Surg*. 2018;107:189-196.
4. Hale DA, Molloy M, Pearl RH, et al. Appendectomy: a contemporary appraisal. *Ann Surg*. 1997;225:252-261.
5. Becker T, Kharbanda A, Bachur R. Atypical clinical features of pediatric appendicitis. *Acad Emerg Med*. 2007;14:124-129.
6. Russell WS, Schuh AM, Hill JG, et al. Clinical practice guidelines for pediatric appendicitis evaluation can decrease computed tomography utilization while maintaining diagnostic accuracy. *Pediatr Emerg Care*. 2013;29:568-573.
7. Bachur RG, Hennelly K, Callahan MJ, et al. Advanced radiologic imaging for pediatric appendicitis, 2005-2009: trends and outcomes. *J Pediatr*. 2012;160:1034-1038.
8. Kobayashi E, Johnson B, Goetz K, et al. Does the implementation of a pediatric appendicitis pathway promoting ultrasound work outside of a children's hospital? *Am J Surg*. 2018;215:917-920.
9. Kharbanda AB, Madhok M, Krause E, et al. Implementation of electronic clinical decision support for pediatric appendicitis. *Pediatrics*. 2016;137:e20151745.
10. Garcia Peña BM, Cook EF, Mandl KD. Selective imaging strategies for the diagnosis of appendicitis in children. *Pediatrics*. 2004;113:24-28.
11. Stephen AE, Segev DL, Ryan DP, et al. The diagnosis of acute appendicitis in a pediatric population: to CT or not to CT. *J Pediatr Surg*. 2003;38:367-371; discussion 367-371.
12. Partrick DA, Janik JE, Janik JS, et al. Increased CT scan utilization does not improve the diagnostic accuracy of appendicitis in children. *J Pediatr Surg*. 2003;38:659-662.
13. Lee SL, Walsh AJ, Ho HS. Computed tomography and ultrasonography do not improve and may delay the diagnosis and treatment of acute appendicitis. *Arch Surg*. 2001;136:556-562.
14. Miglioretti DL, Johnson E, Williams A, et al. The use of computed tomography in pediatrics and the associated radiation exposure and estimated cancer risk. *JAMA Pediatr*. 2013;167:700-707.
15. Brenner D, Elliston C, Hall E, et al. Estimated risks of radiation-induced fatal cancer from pediatric CT. *AJR Am J Roentgenol*. 2001;176:289-296.
16. Kotagal M, Richards MK, Flum DR, et al. Use and accuracy of diagnostic imaging in the evaluation of pediatric appendicitis. *J Pediatr Surg*. 2015;50:642-646.
17. Kotagal M, Richards MK, Chapman T, et al. Improving ultrasound quality to reduce computed tomography use in pediatric appendicitis: the Safe and Sound campaign. *Am J Surg*. 2015;209:896-900; discussion 900.
18. Kharbanda AB, Vazquez-Benitez G, Ballard DW, et al. Development and validation of a novel pediatric appendicitis risk calculator (pARC). *Pediatrics*. 2018;14:e20172699.
19. Bachur RG, Hennelly K, Callahan MJ, et al. Diagnostic imaging and negative appendectomy rates in children: effects of age and gender. *Pediatrics*. 2012;129:877-884.
20. Mandeville K, Pottker T, Bulloch B, et al. Using appendicitis scores in the pediatric ED. *Am J Emerg Med*. 2011;29:972-977.
21. Gausche-Hill M, Schmitz C, Lewis RJ. Pediatric preparedness of US emergency departments: a 2003 survey. *Pediatrics*. 2007;120:1229-1237.
22. Maguire JL, Kulik DM, Laupacis A, et al. Clinical prediction rules for children: a systematic review. *Pediatrics*. 2011;128:e666-e677.
23. Gordon N, Lin T. The Kaiser Permanente Northern California adult member health survey. *Perm J*. 2016;20:34-42.
24. Bornstein S. An integrated EHR at Northern California Kaiser Permanente: pitfalls, challenges, and benefits experienced in transitioning. *Appl Clin Inform*. 2012;3:318-325.
25. Lieu TA, Platt R. Applied research and development in health care: time for a frameshift. *N Engl J Med*. 2017;376:710-713.
26. Ballard DW, Vemula R, Chettipally UK, et al. Optimizing clinical decision support in the electronic health record. Clinical characteristics associated with the use of a decision tool for disposition of ED patients with pulmonary embolism. *Appl Clin Inform*. 2016;7:883-898.
27. Vinson DR, Mark DG, Chettipally UK, et al. Increasing safe outpatient management for emergency department patients with pulmonary embolism: a controlled pragmatic trial. *Ann Intern Med*. 2018;169:855-865.
28. Simon LE, Rauchwerger AS, Chettipally UK, et al. Real-time text message alerts to emergency physicians identifying potential study candidates increases clinical trial enrollment. *J Patient Cent Res Rev*. 2018;5:57.
29. Kaji AH, Schriger D, Green S. Looking through the retrospectoscope: reducing bias in emergency medicine chart review studies. *Ann Emerg Med*. 2014;64:292-298.
30. Samuel M. Pediatric appendicitis score. *J Pediatr Surg*. 2002;37:877-881.
31. Steyerberg EW, Borsboom GJ, van Houwelingen HC, et al. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med*. 2004;23:2567-2586.
32. Altman DG, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:b605.
33. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000;19:453-473.
34. Jamoom E, Hing E. Progress with electronic health record adoption among emergency and outpatient departments: United States, 2006-2011. *NCHS Data Brief*. 2015;(187):1-8.
35. Bennett P, Hardiker NR. The use of computerized clinical decision support systems in emergency care: a substantive review of the literature. *J Am Med Inform Assoc*. 2017;24:655-668.
36. Varghese J, Kleine M, Gessner SI, et al. Effects of computerized decision support system implementations on patient outcomes in inpatient care: a systematic review. *J Am Med Inform Assoc*. 2018;25:593-602.
37. Parikh RB, Kakad M, Bates DW. Integrating predictive analytics into high-value care: the dawn of precision delivery. *JAMA*. 2016;315:651-652.
38. Stiell IG, Wells GA. Methodologic standards for the development of clinical decision rules in emergency medicine. *Ann Emerg Med*. 1999;33:437-447.
39. Almaramhy HH. Acute appendicitis in young children less than 5 years: review article. *Ital J Pediatr*. 2017;43:15.
40. Hamid KA, Mohamed MA, Salih A. Acute appendicitis in young children: a persistent diagnostic challenge for clinicians. *Cureus*. 2018;10:e2347.
41. Altali K, Ruiz-Artacho P, Trenchs V, et al. [Hospital emergency room diagnosis of acute appendicitis in patients aged 2 to 20 years: the INFURG-SEMES score from the emergency infections study of the Spanish Society of Emergency Medicine]. *Emergencias*. 2017;29:231-236.

42. Aydin D, Turan C, Yurtseven A, et al. Integration of radiology and clinical score in pediatric appendicitis. *Pediatr Int*. 2018;60:173-178.
43. Green SM, Schriger DL, Yealy DM. Methodologic standards for interpreting clinical decision rules in emergency medicine: 2014 update. *Ann Emerg Med*. 2014;64:286-291.
44. Benabbas R, Hanna M, Shah J, et al. Diagnostic accuracy of history, physical examination, laboratory tests, and point-of-care ultrasound for pediatric acute appendicitis in the emergency department: a systematic review and meta-analysis. *Acad Emerg Med*. 2017;24:523-551.
45. Bhatt M, Joseph L, Ducharme FM, et al. Prospective validation of the Pediatric Appendicitis Score in a Canadian pediatric emergency department. *Acad Emerg Med*. 2009;16:591-596.
46. Ebell MH, Shinholser J. What are the most clinically useful cutoffs for the Alvarado and Pediatric Appendicitis Scores? a systematic review. *Ann Emerg Med*. 2014;64:365-372.e362.
47. Escriba A, Gamell AM, Fernandez Y, et al. Prospective validation of two systems of classification for the diagnosis of acute appendicitis. *Pediatr Emerg Care*. 2011;27:165-169.
48. Goldman RD, Carter S, Stephens D, et al. Prospective validation of the Pediatric Appendicitis Score. *J Pediatr*. 2008;153:278-282.
49. Pogorelic Z, Rak S, Mrklic I, et al. Prospective validation of Alvarado score and Pediatric Appendicitis Score for the diagnosis of acute appendicitis in children. *Pediatr Emerg Care*. 2015;31:164-168.
50. Dahabreh IJ, Adam GP, Halladay CW, et al. Diagnosis of right lower quadrant pain and suspected acute appendicitis. *AHRQ Comparative Effectiveness Reviews*. 2015;15:EHC025-EF.
51. Kulik DM, Uleryk EM, Maguire JL. Does this child have appendicitis? a systematic review of clinical prediction rules for children with acute abdominal pain. *J Clin Epidemiol*. 2013;66:95-104.
52. Ohmann C, Yang Q, Franke C. Diagnostic scores for acute appendicitis. Abdominal Pain Study Group. *Eur J Surg*. 1995;161:273-281.
53. Carmelli G, Grock A, Picart E, et al. The nitty-gritty of clinical decision rules. *Ann Emerg Med*. 2018;71:711-713.
54. Ekstrom HL, Kharbanda EO, Ballard DW, et al. Development of a clinical decision support system for pediatric abdominal pain in emergency department settings across two health systems within the Healthcare Systems Research Network. *EGEMS (Wash DC)*. 2019;7(1):15.
55. Bendeck SE, Nino-Murcia M, Berry GJ, et al. Imaging for suspected appendicitis: negative appendectomy and perforation rates. *Radiology*. 2002;225:131-136.
56. Kosloske AM, Love CL, Rohrer JE, et al. The diagnosis of appendicitis in children: outcomes of a strategy based on pediatric surgical evaluation. *Pediatrics*. 2004;113(1 pt 1):29-34.
57. Xu J, Adams S, Liu YC, et al. Nonoperative management in children with early acute appendicitis: a systematic review. *J Pediatr Surg*. 2017;52:1409-1415.