UC Davis
UC Davis Previously Published Works

Title

Development and Validation of a Novel Pediatric Appendicitis Risk Calculator (pARC)

Permalink

https://escholarship.org/uc/item/7nh8r5gf

Journal

Pediatrics, 141(4)

ISSN

0031-4005

Authors

Kharbanda, Anupam B
Vazquez-Benitez, Gabriela
Ballard, Dustin W
et al.

Publication Date

2018-04-01

DOI

10.1542/peds.2017-2699

Peer reviewed

Development and Validation of a Novel Pediatric Appendicitis Risk Calculator (pARC)

Anupam B. Kharbanda, MD, MSc,[a] Gabriela Vazquez-Benitez, PhD,[b] Dustin W. Ballard, MD, MBE,[c] David R. Vinson, MD,[c] Uli K. Chettipally, MD, MPH,[c] Mamata V. Kene, MD, MPH,[c] Steven P. Dehmer, PhD,[b] Richard G. Bachur, MD,[d] Peter S. Dayan, MD, MSc,[e] Nathan Kuppermann, MD, MPH,[f] Patrick J. O'Connor, MD, MPH,[b] Elyse O. Kharbanda, MD, MPH[b]

OBJECTIVES: We sought to develop and validate a clinical calculator that can be used to quantify risk for appendicitis on a continuous scale for patients with acute abdominal pain.

METHODS: The pediatric appendicitis risk calculator (pARC) was developed and validated through secondary analyses of 3 distinct cohorts. The derivation sample included visits to 9 pediatric emergency departments between March 2009 and April 2010. The validation sample included visits to a single pediatric emergency department from 2003 to 2004 and 2013 to 2015. Variables evaluated were as follows: age, sex, temperature, nausea and/or vomiting, pain duration, pain location, pain with walking, pain migration, guarding, white blood cell count, and absolute neutrophil count. We used stepwise regression to develop and select the best model. Test performance of the pARC was compared with the Pediatric Appendicitis Score (PAS).

RESULTS: The derivation sample included 2423 children, 40% of whom had appendicitis. The validation sample included 1426 children, 35% of whom had appendicitis. The final pARC model included the following variables: sex, age, duration of pain, guarding, pain migration, maximal tenderness in the right-lower quadrant, and absolute neutrophil count. In the validation sample, the pARC exhibited near perfect calibration and a high degree of discrimination (area under the curve: 0.85; 95% confidence interval: 0.83 to 0.87) and outperformed the PAS (area under the curve: 0.77; 95% confidence interval: 0.75 to 0.80). By using the pARC, almost half of patients in the validation cohort could be accurately classified as at <15% risk or ≥85% risk for appendicitis, whereas only 23% would be identified as having a comparable PAS of <3 or >8.

CONCLUSIONS: In our validation cohort of patients with acute abdominal pain, the pARC accurately quantified risk for appendicitis.

[a]Department of Pediatric Emergency Medicine, Children's Minnesota, Minneapolis, Minnesota; [b]Division of Research, HealthPartners Institute, Bloomington, Minnesota; [c]The Permanente Medical Group, Inc and Division of Research, Kaiser Permanente, Oakland, California; [d]Division of Emergency Medicine, Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts; [e]Division of Pediatric Emergency Medicine, Columbia University Vagelos College of Physicians and Surgeons, New York, New York; and [f]Emergency Medicine and Pediatrics, University of California Davis Health, Sacramento, California

WHAT'S KNOWN ON THIS SUBJECT: Available clinical scores are designed for ease of calculation but have had variable validity and clinical use on external validation. A score that is used to capture more complex interactions between variables may have improved accuracy for predicting appendicitis risk.

WHAT THIS STUDY ADDS: In this derivation and validation study, the pediatric appendicitis risk calculator was used to accurately quantify the risk of appendicitis among children and adolescents presenting to the pediatric emergency department with acute abdominal pain, providing clinically actionable classifications.

Appendicitis remains a common pediatric surgical emergency, with more than 75 000 children diagnosed annually in the United States.[1] Recently, there have been incremental improvements in the evaluation of pediatric patients with possible appendicitis. For example, clinical pathways for acute abdominal pain have revealed the feasibility and effectiveness[2–5] of using ultrasound as first-line imaging without increasing missed diagnoses or negative appendectomy results.[4,5] Reductions in computed tomography (CT) for appendicitis have also been observed in a national sample of 35 pediatric institutions.[6] However, rates for appendiceal perforation have remained unchanged, and variations in care persist.[6]

Clinical scores have been developed to standardize care and limit imaging for patients with possible appendicitis.[7,8] Previous scores were developed with an emphasis on simplicity, to be calculated by hand. Both the Pediatric Appendicitis Score (PAS) and the Alvarado score appeared promising in derivation samples,[7,8] but have revealed variable accuracy and limited clinical use on external validation.[9–11] For example, many patients receive a score signifying intermediate risk (ie, a PAS score of 4–6), encouraging clinicians to seek surgical consultation or advanced diagnostic imaging.[12]

More recently, authors of large observational studies have described variability in the clinical presentation of appendicitis and have emphasized the importance of subgroup analyses and interactions among covariates. For example, age, sex, and duration of symptoms can impact laboratory findings and accuracy of diagnostic imaging in children with acute abdominal pain.[13,14] Age- and sex-specific scores, or scores to identify low-risk patients have been developed by our group and others,[15–17] but these may have limited use in an emergency department (ED) because they cannot be applied to the full range of patients with acute abdominal pain. Widespread adoption of electronic health records[18,19] along with increased use of risk prediction in other domains of medicine[20] provided the impetus to develop a more sophisticated prediction tool for patients with possible appendicitis. Our aim in this investigation was to develop and validate a new pediatric Appendicitis Risk Calculator (pARC), quantifying risk for appendicitis on a continuous scale. Using rigorous methodology for score development and allowing for complex calculations, we aimed for the pARC to have improved accuracy and clinical use over the PAS.

METHODS

Derivation Cohort

We derived the pARC from an existing, deidentified cohort of children with suspected appendicitis.[17] Although the parent study included children 3 to 18.9 years old, given the low risk for appendicitis in children <5 years old and increased likelihood of atypical clinical presentations,[21] the pARC score was derived and validated in patients 5 to 18 years old. In the previous prospective study, conducted from March 2009 through April 2010, ED clinicians collected clinical data from patients with suspected appendicitis at 9 pediatric emergency departments (PEDs). Treating clinicians enrolled children and adolescents who presented to the PED with <96 hours of abdominal pain and who were under evaluation for suspected appendicitis. "Suspected appendicitis" was defined as undergoing laboratory testing, diagnostic imaging, or a surgical consultation for possible appendicitis. Patients with the following conditions were excluded: pregnancy, previous abdominal surgery, inflammatory bowel disease, chronic pancreatitis, sickle cell anemia, cystic fibrosis, a medical condition affecting the ability to obtain an accurate history, or a history of abdominal trauma within the previous 7 days. Study procedures related to training of site staff, patient enrollment, data collection, and data management have been described previously.[17]

Validation Cohort

We validated the pARC using deidentified data from 2 independent cohorts of patients 5 to 18 years old with visits to the Boston Children's Hospital PED from 2003 to 2004 and from 2013 to 2015. These cohorts were chosen as the validation sample because their criteria for cohort entry, data collection, cleaning and quality control were similar to those used in the derivation sample. Consistent with the recommendations of Altman, the validation population did not overlap with the derivation population.[22] Clinical data were collected as part of distinct research[23] and quality improvement projects. For both cohorts, children and adolescents with possible appendicitis were prospectively identified by trained coordinators who screened patients in the PED 10 hours per day. Subjects were included in the cohort when their treating emergency physician ordered advanced imaging or a surgical consult with concern for appendicitis. Parental consent and patient assent were obtained before data collection. Historical, physical examination, and laboratory data were collected in real time. The final diagnosis was based on pathology and surgical reports.

Outcome Measures

In both the derivation and validation cohorts, the primary outcome was appendicitis. For those who underwent an appendectomy, appendicitis was confirmed through the pathology report. To identify missed cases of appendicitis, families

were contacted within 2 to 3 weeks of PED discharge to assess for visits to other sites of care and whether their child had an appendectomy in the interim. For families who could not be contacted, medical records were reviewed for 3 months after the index PED visit. Further description of these methods has been previously published.[15,17,23,24]

Data collection for the parent studies was approved by all participating institutional review boards, with parents or legal guardians consenting to study participation. The current analyses were conducted by using deidentified data sets and were exempt from additional institutional review.

Score Development

Patient history, physical examination, and laboratory variables were collected by using standardized processes in the parent studies.[17,23] We first reviewed distributions, means, medians, ranges, and proportion of variables with missing values in the derivation and validation cohorts.[7,8,17] Predictors evaluated for inclusion in the pARC[7,8,17] were coded as binary variables unless otherwise indicated and included the following: sex, age (5–7.9 years old, 8–11.9 years old, or 12–18 years old for girls and 5–7.9 years old, 8–13.9 years old, or 14–18 years old for boys, accounting for variability in appendicitis risk and alternate causes for acute abdominal pain by age and sex subgroups), a fever in the ED >38°C, duration of pain (<24 hours, 24–47 hours, or 48–96 hours), a history of nausea, a history of emesis, migration of pain to the right-lower quadrant (RLQ), maximum tenderness in the RLQ, abdominal guarding, and pain with walking, coughing, or hopping. For these analyses, "unsure," "don't know," and "missing" responses were coded as not having the sign or symptom. We evaluated the white blood cell (WBC) count and absolute neutrophil count (ANC) as continuous measures (10^3/mL); we assessed for normality and nonlinear associations with appendicitis using a generalized additive model.[25]

For consideration in the pARC, we included only predictors with <10% missing data and at least moderate interrater reliability (κ > 0.35).[24] Because clinical decisions are often made in the context of a patient's age and sex, we evaluated the interactions between age and sex. Following the prognostic model development approach recommend by Royston et al,[26] we selected all potential predictors for the multivariable model on the basis of the following rules: (1) variables associated with appendicitis had a P value <.05 in the age- and sex-adjusted models; (2) associations between variables and appendicitis were in the expected direction; (3) for binary predictors, the β coefficient was >.4; (4) transformation of the laboratory values to a normal scale and shape of the association was informed by graphical exploration; (5) if only the WBC count was available, but not the ANC, the ANC was imputed as ANC = (−0.8783 + 1.1008 × sqrt(WBC))^2; (6) if neither the WBC count nor the ANC was available, the ANC was imputed as 7 × 10^3/μL, corresponding to the mean ANC in our derivation cohort; and (7) interactions between age and sex and each additional predictor with appendicitis were evaluated as potential terms in the model.

Validation

The newly derived pARC was then applied to the validation cohort. On the basis of input from investigators providing clinical care for this population (A.B.K., D.W.B., D.R.V.), we stratified subjects into 7 risk strata as clinically actionable categories: <5%, 5% to 14%, 15% to 24%, 25% to 49%, 50% to 74%, 75% to 84%, and ≥85%. We evaluated calibration of the pARC in the validation cohort by plotting observed and predicted risks and used the Hosmer and Lemeshow[27] goodness of fit test on the basis of a decile partition. We evaluated the discriminatory performance using the area under the curve (AUC) plot and AUC statistic.

For the validation cohort, we compared the calibration and discrimination performance of the pARC versus the previously published PAS. We also evaluated clinical use of the PAS and pARC, comparing the proportion of subjects in the validation cohort classified as high risk or low risk for appendicitis.

Exploratory Analyses

We conducted additional exploratory analyses related to potential clinical applications of the pARC score. Using the larger derivation sample, we evaluated positive, negative, and equivocal ultrasound findings, stratified by pARC score.

RESULTS

Study Population

The derivation cohort included 2423 children and adolescents, 40% of whom had appendicitis. The validation cohort had 1426 patients, 35% of whom had appendicitis. Clinical characteristics of the 2 cohorts are presented in Table 1. Derivation and validation subjects were similar in age and sex. Reported rates of nausea or vomiting, pain with walking or hopping, and migration of pain to the RLQ were all higher in the derivation cohort. Similarly, on physical examination, maximal tenderness in the RLQ and guarding were more common in the derivation cohort.

Score Development

All predictors evaluated in the derivation set were significantly associated with appendicitis risk (P value < .05), adjusted by age and sex (Table 2). Fever was missing for

TABLE 1 Characteristics of Derivation and Validation Cohorts

	Derivation Cohort, $N = 2423$	Validation Cohort, $N = 1426$
No. PEDs	9	1
Years included	2009–2010	2003–2004, 2013–2015
Median age, y (IQR)	11 (8–14)	11 (8–14)
Sex and age, *n* (%)		
Male	1217 (50)	686 (48)
5–7.9 y	256 (11)	139 (10)
8–13.9 y	708 (29)	319 (22)
14–18 y	253 (10)	228 (16)
Female	1206 (50)	740 (52)
5–7.9 y	205 (8)	128 (9)
8–11.9 y	419 (17)	283 (20)
12–18.9 y	582 (24)	329 (23)
Reported clinical presentation, *n* (%)		
Duration of abdominal pain, h		
<24	1344 (55)	724 (51)
24–47	587 (24)	438 (31)
48–96	492 (20)	264 (19)
Fever in ED	440 (18)	275 (19)
Nausea or vomiting	1689 (70)	845 (59)
Pain with walking or hopping	1694 (70)	785 (55)
Migration of pain to RLQ	1035 (43)	474 (33)
Findings on examination, *n* (%)		
Maximal tenderness in RLQ	1691 (70)	838 (59)
Abdominal guarding	1339 (55)	592 (42)
Median laboratory results (IQR)		
WBC ($\times 10^3$/μL)	11.2 (7.4–15.9)	10.0 (6.9–14.6)
ANC ($\times 10^3$/μL)	7.9 (3.9–12.6)	7.0 (3.7–11.8)
Appendicitis confirmed, *n* (%)	968 (40)	493 (35)

Missing values are set as not having the clinical characteristic. *N* (%) is for the clinical characteristic being present. IQR, interquartile range.

TABLE 2 Analysis of Individual Predictors of Appendicitis (Derivation Set: $n = 2423$)

Clinical Characteristic	Missing Values, *n* (%)	PPV (*P*)	β[a] (95% CI)	*P*
Duration of pain, 24–47 h	0 (0)	0.47 (<.0001)	.30 (0.10 to 0.50); Reference <24 h	.003
Fever	440 (18)	0.45 (.02)	.26 (0.04 to 0.48)	.02
Nausea or vomiting	51 (2)	0.44 (<.0001)	.59 (0.41 to 0.78)	<.0001
Pain with walking, hopping, or coughing	177 (7)	0.48 (<.0001)	1.3 (1.0 to 1.8)	<.0001
Migration of pain to RLQ	152 (6)	0.51 (<.0001)	1.3 (1.1 to 1.5)	<.0001
Maximal tenderness in RLQ	96 (4)	0.48 (<.0001)	1.2 (1.0 to 1.4)	<.0001
Abdominal guarding	22 (1)	0.51 (<.0001)	1.05 (0.87 to 1.2)	<.0001
WBC count ($\times 10^3$/μL)	172 (7)	—	—	—
ANC ($\times 10^3$/μL)	216 (9)	—	—	—

—, not applicable.

[a] The model included the clinical characteristic and sex-age covariates; for PPV and β coefficient, missing values were set as not having the clinical characteristic.

18% in the derivation cohort and so was not entered in the pARC model. The relationship between WBC count or ANC and appendicitis risk did not vary across 6 predefined age and sex subgroups. However, the relationship between WBC count, ANC and risk for appendicitis was linear up to a threshold of 20×10^3/μL and 14×10^3/μL,

respectively. Beyond these thresholds, increases in WBC count or ANC were not associated with increases in appendicitis risk. We used this information to model WBC count and ANC as a 2-step linear function (Supplemental Fig 4).

An initial appendicitis risk model was developed with the remaining categorical predictors (duration of pain, combination of nausea or vomiting, pain with walking, migration of pain to RLQ, maximal tenderness in the RLQ, guarding) and with the ANC as a linear function up to a threshold of 14×10^3/μL and a constant function for higher values. In the initial model, the combination of nausea or vomiting was not significant (β coefficient .16; *P* value = .18), so it was not included in the final model. In Table 3, we present the final pARC model with the associated β coefficients. The concordance statistic for the final model was 0.86 (95% confidence interval [CI]: 0.85 to 0.88).

Score Validation

Complete data for validation of the pARC were available for 1426 patients. Across 7 clinically actionable risk categories (<5%, 5%–14%, 15%–24%, 25%–49%, 50%–74%, 75%–84%, and ≥85%), the pARC score provided valid risk prediction. The AUC for the pARC was 0.85 (95% CI: 0.83 to 0.87) (Table 4).

Score Comparison

To compare the PAS and pARC, we developed PAS to pARC conversions using appendicitis rates by individual PAS from our validation cohort and compared results to those in previous studies[4,9,11,12]: <2: <5%; 2: 5% to 14%; 3: 15% to 24%; 4–5: 25% to 49%; 6: 50% to 74%; 7–8: 75% to 84%; 9–10: ≥85%. The calibration plot for the pARC and PAS is demonstrated in Fig 1. Both scores predicted appendicitis with high accuracy. In Fig 2, we demonstrate that in the validation cohort, the pARC score had an AUC greater than that for the PAS (0.85 [95% CI: 0.83 to 0.87] versus 0.77 [95% CI: 0.75 to 0.80], respectively). The use of the pARC also allowed more subjects to be classified as ≥85% risk or <15%

TABLE 3 Risk Score Equation From Logistic Regression Analysis

Clinical Characteristic	Final Model, β Coefficient (95% CI)	P
Intercept	−8.7	NA
Male sex	1.28 (0.89 to 1.66)	<.0001
Age and sex		
Age 5–7.9 y	.38 (−0.04 to 0.80)	.08
Male 5–7.9 y	−1.05 (−1.65 to −0.44)	.001
Age 8–13.9 y	−.72 (−1.21 to −0.23)	.004
Male 14–18 y, Female 12–18 y	Reference	NA
Duration of pain, h		
<24	Reference	NA
24– <48	.47 (0.22 to 0.72)	.001
48–96	.10 (−0.18 to 0.38)	.49
Presence of pain with walking	1.05 (0.80 to 1.30)	<.0001
History of migration of pain to RLQ	.46 (0.24 to 0.67)	<.0001
Maximal tenderness in RLQ	1.14 (0.89 to 1.40)	<.0001
Abdominal guarding	.67 (0.46 to 0.89)	<.0001
For ANC <14 × 10^3/μL, $\sqrt{ANC}$	1.77 (1.56 to 1.99)	<.0001
For ANC ≥14 × 10^3/μL, constant	6.62 (5.94 to 7.29)	<.0001

The concordance statistic for the final model was 0.864. NA, not applicable.

risk for appendicitis, as compared with PAS (Fig 3). Even in patients with a PAS >8, the maximum positive predictive value (PPV) was 81%. Full test performance of the PAS is provided in Supplemental Table 6.

The pARC and Ultrasound

In Table 5, we present the relationship between the pARC strata, ultrasound use, and performance of ultrasound. Of 2423 subjects in our derivation sample, 905 (37%) had an ultrasound, with use being the highest in the 2 lowest pARC strata (45% for pARC scores of <5% risk and 46% for pARC scores of 5% to 14% risk). Overall, 443 (49%) of patients who underwent an ultrasound had an equivocal study, and the rate

of appendicitis after an equivocal ultrasound was 18%. The PPV of ultrasound was ≤70% for patients with a pARC score of <25%. For pARC risk strata of 25% or higher, the PPV of ultrasound increased to ≥94%.

DISCUSSION

In this large multicenter study, we have demonstrated that the newly derived and validated pARC can be used to provide an accurate and discrete assessment of a patient's risk for appendicitis, with improved accuracy and clinical use as compared with a previously published appendicitis score. Importantly, in our validation cohort with a background risk of appendicitis of 35%, the pARC score was able to classify half of patients

as at ≥85% risk or <15% risk for appendicitis, thresholds where surgical evaluation or observation, respectively, may be recommended over immediate diagnostic imaging.

Advanced diagnostic imaging remains common for patients with suspected appendicitis,[28] with upwards of 80% undergoing ultrasound, CT, or MRI.[29] Kotagal et al[30] recently reported that among 2538 children with an appendectomy in the Washington State surgical database from 2008 to 2012, 99.7% had a CT or ultrasound before surgery. Equally concerning, among patients with undifferentiated abdominal pain, Fahimi et al[31] described that within the National Hospital Ambulatory Medical Center Survey, 52% underwent imaging with CT and 25% underwent imaging with CT or ultrasound. Despite the high use, the type of diagnostic imaging has shifted dramatically over the past decade, with Bachur et al[6] reporting a 48% decline in CT use from 2010 to 2013 for patients with appendicitis across 35 children's hospitals. During this period, the use of ultrasound increased 46%, and overall imaging rates remained unchanged.[6]

The decline in use of CT is significant because fewer children are exposed to ionizing radiation.[32] However, the high use of ultrasound raises the potential for overuse. The

TABLE 4 Predicted and Observed Rate of Appendicitis and Performance of the pARC at Each Score Cut Point in the Validation Sample

Predicted Risk for Appendicitis		Observed Rate of Appendicitis, %	Proportion of Patients Identified, n (%)	PPV, %	Sensitivity, %	Specificity, %	Missed Appendicitis Rate, %	Negative Appendectomy Results Rate, %
Low	<5%	5	282 (20)	34.6	100.0	0.0	0.4	8.8
	5–14%	10	313 (22)	41.9	97.2	28.7	0.4	7.7
Intermediate	15–24%	23	167 (12)	53.7	90.7	58.7	0.6	6.8
	25–49%	42	262 (18)	61.4	82.8	72.5	0.8	5.2
	50–74%	62	221 (16)	73.7	60.2	88.8	1.2	5.5
	75–84%	78	91 (6)	87.4	32.3	97.5	1.1	2.6
High	≥85%	97	91 (6)	96.7	17.8	99.7	0	1.2

Validation sample (n = 1426; 35% with appendicitis); test result was positive if the score was greater than or equal to the cut point.
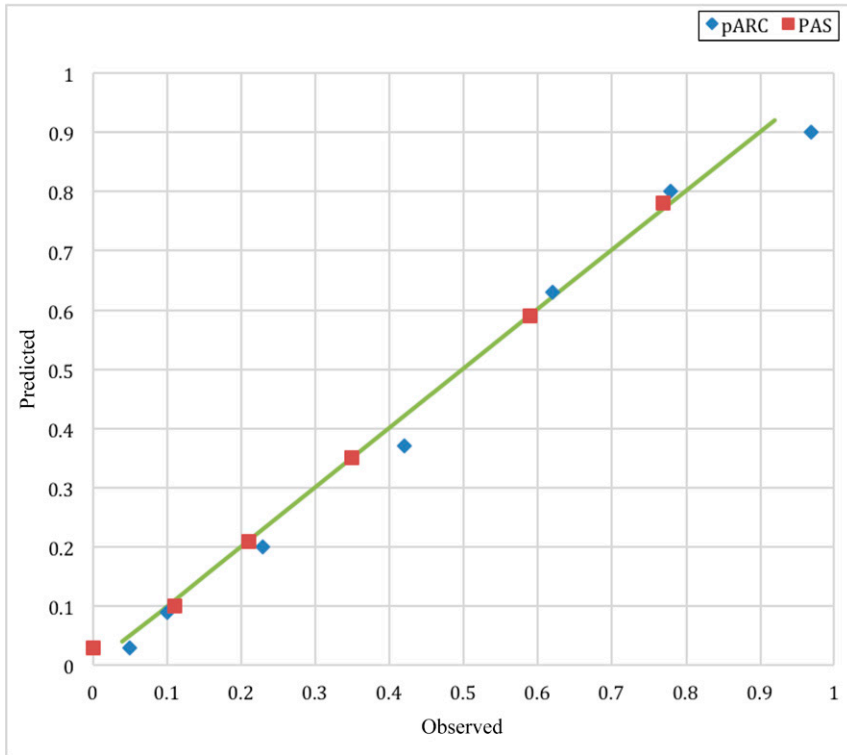
FIGURE 1
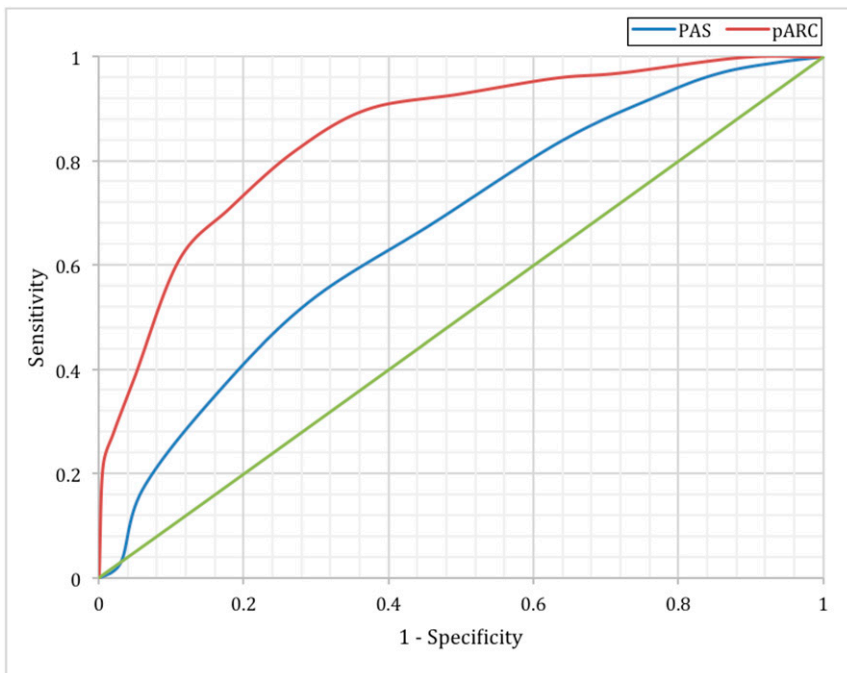Calibration plot of the pARC and PAS comparison (*n* = 1426).

FIGURE 2
Comparison of the pARC and PAS in the validation sample (*n* = 1426). The AUC for the PAS was 0.77 (95% CI: 0.75 to 0.80). The AUC for the pARC was 0.85 (95% CI: 0.83 to 0.87).

unintended consequences of increased use of ultrasound may be increases in ED length of stay, hospital expenditures, and false-positive or indeterminate study results.[29] The most impactful consequence of the increased ultrasound may be the likelihood of nonvisualization of the appendix and thus an equivocal interpretation. In some centers, half of ultrasounds are reported as equivocal[33]; these equivocal ultrasounds may compel providers to order a CT, MRI, or admit for observation.[34] This is especially concerning in cases in which the a priori risk for appendicitis is low and highlights the need for judicious, risk-stratified use of any diagnostic imaging.

In previous studies, appendicitis scores were touted as mechanisms to decrease the use of CT and standardize care for patients with acute abdominal pain. In the 2 most commonly cited scores, derived by Alvarado[8] and Samuel,[7] the authors assign point values to patient history, physical examination, and laboratory findings. Points are summed and cutoffs applied to define low, intermediate, and high risk groups. These scores are easy to calculate, but upwards of 70% of patients may be assigned scores that do not aid in diagnostic assessment (ie, risk scores of 4, 5, or 6).[9,12,35,36] The high proportion of patients assigned scores in which appendicitis can neither be ruled in nor ruled out has also been demonstrated in several real world implementation studies.[3,37] For example, Depinet et al[4] found that 61% of 489 pediatric patients with acute abdominal pain at a single center received a PAS score between 3 and 6, which the authors classified as equivocal or medium risk.

The use of diagnostic imaging for patients with acute abdominal pain may be magnified when patients are assigned intermediate appendicitis risk scores. It is in this context that the pARC may be most impactful on clinical care. In our study, up to 40% of the pediatric patients who presented with acute abdominal pain would have been assigned a low risk pARC score by which the clinician
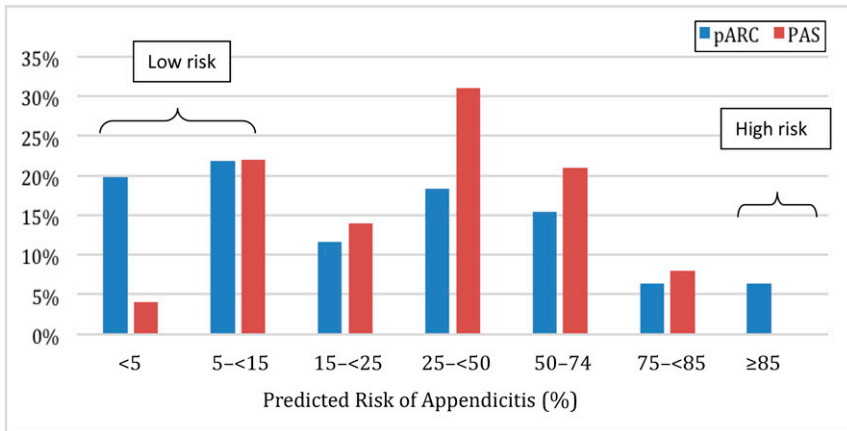
FIGURE 3
Comparison of the pARC and PAS based on predicted risk of appendicitis in the validation sample (*n* = 1426).

could defer diagnostic imaging. Furthermore, in the subset assigned a pARC score of 15% to 24%, the high rate of equivocal ultrasound readings and ultimate low risk for appendicitis suggests that these patients may be managed with observation rather than ultrasound. Finally, among patients with pARC scores ≥85%, it would be reasonable and safe to encourage surgical evaluation with selective use of diagnostic imaging. The potential implications of this approach merit discussion among multidisciplinary care teams. Suggested in the data presented here is that broad use of the pARC score at the point of care could facilitate a reduction in the use of ultrasound, CT, and MRI.

One strength of our study lies in the approach used to develop the pARC, because it differed substantially from methods used in the derivation of previous appendicitis risk scores. As outlined by Royston et al,[26] we selected clinically relevant candidate predictors, evaluated the quality of the data, developed a strategy to consistently model continuous variables, identified the influence of outliers, and considered multiple potential interactions between predictors and impact on overall model performance. Next, for rule validation, we followed the guidance outlined by Altman et al,[22] in that we validated our model on a similar but not contemporaneous patient population. Because the aim of most authors of prognostic

studies is to create clinically valuable risk scores or indexes, the definition of risk groups should be driven by clinical rather than statistical criteria. Strengths of the pARC include its validity in predicting appendicitis risk and its classification of patients into clinically actionable risk groupings. Finally, we selected our model to favor a minimal number of predictors, applying a priori knowledge regarding reproducibility of predictors.[24] More complex models are prone to overfitting data, with little practical gain.[26]

Several limitations should be noted. First, a few variables we considered for inclusion had substantial missing data and could not be incorporated in the model. Second, the pARC was derived and validated by using data from patients at children's hospitals. Furthermore, for the validation cohort, we aggregated data from different time periods from a single children's hospital. As such, our results require validation in new populations before widespread dissemination. Similarly, our derivation and validation cohorts had appendicitis at rates of 40% and 35%, respectively. The discriminative power of the pARC may be diminished if applied in populations with higher or lower appendicitis rates. In addition,

TABLE 5 Application of the pARC for Evaluating Ultrasound Use Within the Derivation Sample

pARC Score Strata, %	Ultrasound Performed, *n* (%)	Appendicitis in Those With Ultrasound Performed, *n* (%)	Equivocal Ultrasound, *n* (%)	Equivocal Ultrasound With Appendicitis, *n* (%)	NPV of Ultrasound, %	PPV of Ultrasound, %
<5	148 (45)	9 (6.1)	89 (60)	5 (6)	100	67
5–14	188 (46)	18 (9.6)	109 (58)	3 (3)	97	68
15–24	102 (38)	12 (11.8)	59 (58)	4 (7)	97	70
25–49	174 (36)	59 (33.9)	85 (49)	15 (18)	98	94
50–74	153 (34)	98 (64.1)	65 (42)	33 (51)	100	97
75–84	85 (35)	66 (77.6)	21 (25)	11 (52)	67	98
≥85	55 (23)	49 (89.1)	15 (27)	9 (60)	—	100
Overall	905 (37)	311 (34.4)	443 (49)	80 (18)	96	93

The derivation sample total was 2423. Ultrasounds were performed for 905 patients, and appendicitis occurred in 311 patients. NPV was used to indicate a negative test result for appendicitis and/or a negative ultrasound test result. PPV was used to indicate a positive test result for appendicitis and/or a positive ultrasound test result. Equivocal ultrasounds were not included in the calculation of NPV and PPV. NPV, negative predictive value; —, not applicable.

the pARC is not intuitive and requires sophisticated calculations. Nevertheless, the pARC can be easily programmed and integrated within the electronic health record, promoting meaningful use of available clinical and laboratory data.

CONCLUSIONS

In this derivation and validation study, the pARC was used to accurately quantify the risk of appendicitis among children and adolescents presenting to the PED with acute abdominal pain. Next steps include a prospective validation of the pARC and an evaluation of how the pARC may impact the care delivered.

ACKNOWLEDGMENTS

We thank the investigators who assisted in data collection at each of the enrolling sites from which the pARC was developed and validated.

ABBREVIATIONS

ANC: absolute neutrophil count
AUC: area under the curve
CI: confidence interval
CT: computed tomography
ED: emergency department
pARC: pediatric Appendicitis Risk Calculator
PAS: Pediatric Appendicitis Score
PED: pediatric emergency department
PPV: positive predictive value
RLQ: right-lower quadrant
WBC: white blood cell

REFERENCES

1. Barrett ML, Hines AL, Andrews RM. *Trends in Rates of Perforated Appendix, 2001-2010: Healthcare Cost and Utilization Project (HCUP) Statistical Brief #159.* Rockville, MD: Agency for Healthcare Research and Quality; 2013

2. Shah SR, Sinclair KA, Theut SB, Johnson KM, Holcomb GW III, St Peter SD. Computed tomography utilization for the diagnosis of acute appendicitis in children decreases with a diagnostic algorithm. *Ann Surg.* 2016;264(3):474–481

3. Santillanes G, Simms S, Gausche-Hill M, et al. Prospective evaluation of a clinical practice guideline for diagnosis of appendicitis in children. *Acad Emerg Med.* 2012;19(8):886–893

4. Depinet H, von Allmen D, Towbin A, Hornung R, Ho M, Alessandrini E. Risk stratification to decrease unnecessary diagnostic imaging for acute appendicitis. *Pediatrics.* 2016;138(3):e20154031

5. Kharbanda AB, Madhok M, Krause E, et al. Implementation of electronic clinical decision support for pediatric appendicitis. *Pediatrics.* 2016;137(5):e20151745

6. Bachur RG, Levy JA, Callahan MJ, Rangel SJ, Monuteaux MC. Effect of reduction in the use of computed tomography on clinical outcomes of appendicitis. *JAMA Pediatr.* 2015;169(8):755–760

7. Samuel M. Pediatric appendicitis score. *J Pediatr Surg.* 2002;37(6):877–881

8. Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med.* 1986;15(5):557–564

9. Schneider C, Kharbanda A, Bachur R. Evaluating appendicitis scoring systems using a prospective pediatric cohort. *Ann Emerg Med.* 2007;49(6):778–784, 784.e1

10. 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(1):95–104

11. 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(4):365–372.e2

12. Bhatt M, Joseph L, Ducharme FM, Dougherty G, McGillivray D. Prospective validation of the pediatric appendicitis score in a Canadian pediatric emergency department. *Acad Emerg Med*. 2009;16(7):591–596

13. Bachur RG, Dayan PS, Dudley NC, et al. The influence of age on the diagnostic performance of white blood cell count and absolute neutrophil count in suspected pediatric appendicitis. *Acad Emerg Med*. 2016;23(11):1235–1242

14. Bachur RG, Dayan PS, Bajaj L, et al; Pediatric Emergency Medicine Collaborative Research Committee of the American Academy of Pediatrics. The effect of abdominal pain duration on the accuracy of diagnostic imaging for pediatric appendicitis. *Ann Emerg Med*. 2012;60(5):582–590.e3

15. Kharbanda AB, Monuteaux MC, Bachur RG, et al; Pediatric Emergency Medicine Collaborative Research Committee of the American Academy of Pediatrics. A clinical score to predict appendicitis in older male children. *Acad Pediatr*. 2017;17(3):261–266

16. Scheller RL, Depinet HE, Ho ML, Hornung RW, Reed JL. Utility of pediatric appendicitis score in female adolescent patients. *Acad Emerg Med*. 2016;23(5):610–615

17. Kharbanda AB, Dudley NC, Bajaj L, et al; Pediatric Emergency Medicine Collaborative Research Committee of the American Academy of Pediatrics. Validation and refinement of a prediction rule to identify children at low risk for acute appendicitis. *Arch Pediatr Adolesc Med*. 2012;166(8):738–744

18. Nakamura MM, Harper MB, Jha AK. Change in adoption of electronic health records by US children's hospitals. *Pediatrics*. 2013;131(5). Available at: www.pediatrics.org/cgi/content/full/131/5/e1563

19. 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

20. Dziadzko MA, Gajic O, Pickering BW, Herasevich V. Clinical calculators in hospital medicine: availability, classification, and needs. *Comput Methods Programs Biomed*. 2016;133:1–6

21. Becker T, Kharbanda A, Bachur R. Atypical clinical features of pediatric appendicitis. *Acad Emerg Med*. 2007;14(2):124–129

22. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:b605

23. Kharbanda AB, Taylor GA, Fishman SJ, Bachur RG. A clinical decision rule to identify children at low risk for appendicitis. *Pediatrics*. 2005;116(3):709–716

24. Kharbanda AB, Stevenson MD, Macias CG, et al; Pediatric Emergency Medicine Collaborative Research Committee of the American Academy of Pediatrics. Interrater reliability of clinical findings in children with possible appendicitis. *Pediatrics*. 2012;129(4):695–700

25. Hastie TJ, Tibshirani RJ. *Monographs on Statistics and Applied Probability 43: Generalized Additive Models*. Boca Raton, FL: Chapman & Hall/CRC; 1990

26. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ*. 2009;338:b604

27. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. 2nd ed. New York, NY: John Wiley & Sons, Inc; 2000

28. Parker MW, Shah SS, Hall M, Fieldston ES, Coley BD, Morse RB. Computed tomography and shifts to alternate imaging modalities in hospitalized children. *Pediatrics*. 2015;136(3). Available at: www.pediatrics.org/cgi/content/full/136/3/e573

29. Bachur RG, Hennelly K, Callahan MJ, Monuteaux MC. Advanced radiologic imaging for pediatric appendicitis, 2005-2009: trends and outcomes. *J Pediatr*. 2012;160(6):1034–1038

30. Kotagal M, Richards MK, Flum DR, Acierno SP, Weinsheimer RL, Goldin AB. Use and accuracy of diagnostic imaging in the evaluation of pediatric appendicitis. *J Pediatr Surg*. 2015;50(4):642–646

31. Fahimi J, Herring A, Harries A, Gonzales R, Alter H. Computed tomography use among children presenting to emergency departments with abdominal pain. *Pediatrics*. 2012;130(5). Available at: www.pediatrics.org/cgi/content/full/130/5/e1069

32. 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(8):700–707

33. Larson DB, Trout AT, Fierke SR, Towbin AJ. Improvement in diagnostic accuracy of ultrasound of the pediatric appendix through the use of equivocal interpretive categories. *AJR Am J Roentgenol*. 2015;204(4):849–856

34. Bachur RG, Rangel SJ. The threat of diagnostic uncertainty in the medical management of uncomplicated appendicitis. *JAMA Pediatr*. 2017;171(6):505–506

35. Meltzer AC, Baumann BM, Chen EH, Shofer FS, Mills AM. Poor sensitivity of a modified Alvarado score in adults with suspected appendicitis. *Ann Emerg Med*. 2013;62(2):126–131

36. Goldman RD, Carter S, Stephens D, Antoon R, Mounstephen W, Langer JC. Prospective validation of the pediatric appendicitis score. *J Pediatr*. 2008;153(2):278–282

37. Saucier A, Huang EY, Emeremni CA, Pershad J. Prospective evaluation of a clinical pathway for suspected appendicitis. *Pediatrics*. 2014;133(1). Available at: www.pediatrics.org/cgi/content/full/133/1/e88