

Ultrasonographic findings of child acute appendicitis incorporated into a scoring system

Gheorghe Nicusor Pop^{1,2}, MD, Flaviu Octavian Costea¹, MD, Diana Lungeanu², PhD, Emil Radu Iacob^{3,4}, MD, PhD, Calin Marius Popoiu^{3,4}, MD, PhD

INTRODUCTION This study aimed to investigate the ultrasonographic features of paediatric acute appendicitis and incorporate them into a scoring algorithm that will quantify the risk of complications and the strength of recommendation for surgical intervention.

METHODS 179 patients with suspected appendicitis who had undergone ultrasonographic examination were included in this study. Based on their medical evaluation and post-surgical histopathological results, patients were categorised into confirmed appendicitis (n = 101) and non-appendicitis (n = 78) groups.

RESULTS In the appendicitis group, the appendix was visualised in 66 (65.3%) patients. In cases where the appendix was not visualised, we looked out for secondary inflammatory signs, which were present in 32 (31.7%) patients. Using stepwise logistic regression, Blumberg's sign, free fluid or collection, hyperaemia, non-compressible appendix and an appendix diameter > 7 mm were found to be significant predictive factors for appendicitis. A new scoring system called POPs was developed, combining inflammatory predictors and ultrasonography findings, with an area under the receiver operating characteristic curve of 0.958 (95% confidence interval 0.929–0.986).

CONCLUSION The newly developed POPs-based diagnosis scheme proved a promising alternative to existing scoring systems such as the Alvarado score. Although further calibration would be beneficial, the proposed scoring scheme is simple and easy to understand, memorise and apply in the emergency room.

Keywords: acute appendicitis, child, diagnostic ultrasound, risk assessment

INTRODUCTION

Acute appendicitis is a common disease in children, and appendectomy is the most common emergency operation performed in this population.⁽¹⁾ The clinical diagnosis of acute appendicitis remains difficult owing to various confounders: the presentation is often atypical; the child has difficulties in expressing the symptoms, and the overall clinical picture is similar to other paediatric pathologies. Preoperative diagnosis of acute appendicitis in children results in a high rate of negative appendectomies, generally accepted to minimise the risk of perforations. For example, recent studies reported negative appendectomy rates between 3% and 11%.^(2,3)

Several scoring systems have been designed as an alternative to or supplement in diagnosing child acute appendicitis: the Alvarado score;⁽⁴⁾ Paediatric Appendicitis Score;⁽⁵⁾ Appendicitis Inflammatory Response score;⁽⁶⁾ and Children's Appendicitis Score.⁽⁷⁾ All these scoring systems aim at reducing the negative appendectomy rate without increasing the proportion of perforations, as children with uncomplicated appendicitis can benefit from adequate non-surgical management.

It is difficult to make an accurate diagnosis of acute appendicitis purely based on signs and symptoms, and thus, ultrasonography (US) is a valuable supplementary modality. In 1986, Puylaert was the first person to introduce the graded compression technique for the diagnosis of acute appendicitis.⁽⁸⁾ Although US technology has evolved considerably over the years, the diagnosis of acute

appendicitis through this method has undergone few changes. Despite widely varying in its reported accuracy, the use of US in children is valuable because it considerably reduces the use of computed tomography (CT), which involves irradiation and high costs. For example, in a study of 2,180 children suspected of having acute appendicitis, Dibble et al reported 98.7% sensitivity and 97.1% specificity for US examinations.⁽⁹⁾ However, a key component of US is its dependency on the technical skill of the operator: an accurate US examination is directly correlated with the operator's training and experience. US examination of the abdomen in suspicious appendicitis also gives excellent diagnostic value for differential diagnoses of acute appendicitis and other more frequent inflammatory diseases of the ileocecal region (e.g. mesenteric lymphadenitis, ovarian cyst).⁽¹⁰⁾

Shogilev et al were the first to report that although laboratory markers had limited diagnostic utility on their own, they were worthwhile when used in combination.⁽¹¹⁾ Diagnostic values of white blood count (WBC), neutrophil percentages and C-reactive protein (CRP) were proven to be helpful in other studies,⁽⁷⁾ and the presence of ketone bodies in urine was also reported as important for clinical decision-making in patients with clinically suspected appendicitis.⁽¹²⁾ Moreover, an enlarged diameter of the appendix is considered as an invaluable predictive factor for appendicitis.⁽¹³⁾

Although US investigation has been widely used in the diagnosis of appendicitis as a complement to existing scoring systems, there is currently no scoring system that incorporates

¹Victor Babeş University of Medicine and Pharmacy, ²Centre for Modelling Biological Systems and Data Analysis, Department of Functional Sciences, ³Department of Pediatric Surgery, Victor Babeş University of Medicine and Pharmacy, ⁴The Louis Turcanu Emergency Clinical Hospital for Children, Timișoara, Romania

Correspondence: Dr Flaviu Octavian Costea, Consultant, Victor Babeş University of Medicine and Pharmacy, Piata Eftimie Murgu 2, Timișoara, 300041, Romania. tavi.costea90@gmail.com

US findings. The aim of this study was to evaluate the US features of acute appendicitis in children and transpose them into a scoring algorithm to be used in the emergency setting.

METHODS

The study followed a retrospective cross-sectional design; it was carried out on routinely collected medical data in Victor Babeş University of Medicine and Pharmacy, Romania, and approved by the hospital's ethics committee. The hospital is a teaching facility affiliated to the University of Medicine and Pharmacy. Upon admittance, the parents of all paediatric patients provided written informed consent for secondary use of medical data on the condition of prior de-identification. Furthermore, patients who were able to understand the aims of the study (based on their age, maturity and condition) were given the opportunity to decide if they wanted to contribute their de-identified data to medical research by signing an informed consent form or decline if they were not interested. Taking all these into consideration, no patient or parental supplementary consent was required for this secondary use of medical data.

The medical records, spanning a 12-month period between January and December 2017, were reviewed and all suspected appendicitis cases were initially included in the study, giving a total of 224 patients. Of these, 45 patients who did not undergo US examinations were excluded. Among the 179 remaining patients, 109 appendectomies were performed. Based on the histopathological outcome of the appendectomy specimens as the gold standard for diagnosis, the patients were categorised into two distinct groups: an appendicitis group ($n = 101$) with confirmed acute appendicitis and a non-appendicitis group ($n = 78$) with other pathologies (e.g. mesenteric lymphadenitis, ovarian cyst, Meckel diverticulitis). Fig. 1 shows the algorithm for patients who were included in the study.

Data collected for each patient included age, gender, geographical area of residence (either rural or urban), general symptoms (e.g. duration, characteristics, localisation and migration of pain; pain aggravation by moving, coughing or walking; anorexia; nausea; vomiting; fever), results of physical examination (e.g. localised tenderness, rebound tenderness, pain on percussion/coughing, generalised guarding, bowel sound characteristics), laboratory test results (e.g. WBC counts, CRP level and presence of urinary ketone bodies presence), surgical procedure and complications, antibiotic use, US findings, Alvarado score and histopathological outcome.

US examinations were performed by the radiologist on call using a LOGIQ-e (GE Healthcare, Wauwatosa, WI, USA) US machine with a 4–7 MHz convex transducer, followed by investigation of the lower right abdomen with the gradual compression technique using a 12-MHz linear transducer. The following clinical findings were obtained from the patients' medical records: appendix diameter; degree of compressibility; visualised appendicoliths; oedema/hyperaemia; vascularisation; inflammation of the adjacent fat; free liquid in the pouch of Douglas; and mesenteric lymphadenopathy.

Continuous variables were presented as mean and standard

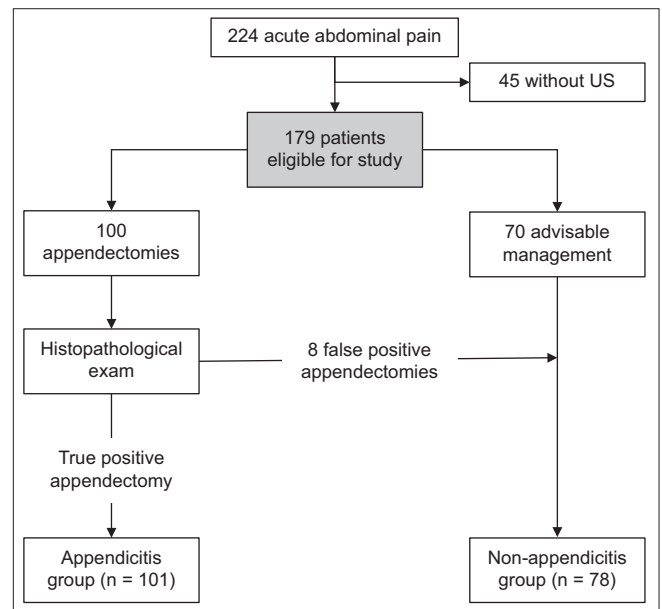


Fig. 1 Chart shows the algorithm of the patients included in the study.

deviation or median and interquartile range based on their distribution (i.e. either normal or otherwise, respectively). Categorical data was described as frequency counts, and their respective percentages were calculated from the total. Descriptive and inferential statistical analysis was performed to summarise the characteristics of the study population. Chi-square or Fisher's exact test was used to evaluate the significance of differences in the proportions of clinical and US findings. Student's t -test and Mann-Whitney U test were used for comparing continuous variables based on their distribution. The receiver operating characteristic (ROC) curve was used to illustrate diagnostic ability, and the thresholds to discriminate between the two groups were determined using the Youden index. DeLong test⁽¹⁴⁾ was used to compare the areas under the ROC curves (e.g. the Alvarado score and our proposed scoring). Stepwise logistic regression and the Akaike information criterion were used in the analysis of the contributing/predictive factors in diagnostic decision-making and appendicitis scoring scheme. A p -value < 0.05 was considered statistically significant. Data analysis was performed using IBM SPSS Statistics version 25.0 (IBM Corp, Armonk, NY, USA).

RESULTS

A total of 179 patients were included in the study. The patients were aged 2–17 (mean 10.32 ± 3.69) years, with 107 (59.8%) male patients. Their demographics, clinical characteristics and laboratory results are presented in Table I. There were no statistically significant differences between the appendicitis and non-appendicitis groups in the distribution of age ($p = 0.785$) and gender ($p = 0.265$). The duration of hospitalisation was 8.7 ± 3.0 days and 3.8 ± 1.8 days for the appendicitis and non-appendicitis groups, respectively ($p < 0.001$).

Of the 109 appendectomies performed, the laparoscopic surgery technique was used in 26 (23.9%) interventions and the classical approach in 75 (68.8%) interventions. Antibiotogram was positive in 34 (31.2%) of the surgical cases, and *Escherichia*

Table I. Demographics, clinical characteristics and laboratory results of patients in the appendicitis and non-appendicitis groups.

Characteristic	No. (%)		p-value
	Appendicitis (n = 101)	Non-appendicitis (n = 78)	
Age* (yr)	10.39 ± 3.81	10.24 ± 3.56	0.785
Male gender	64 (63.4)	43 (55.1)	0.265
Reside in urban area	51 (50.5)	41 (52.6)	0.748
Duration of hospitalisation* (day)	8.72 ± 3.06	3.82 ± 1.82	< 0.001
RLQ tenderness	97 (96.0)	33 (42.3)	< 0.001
Blumberg's sign	66 (65.3)	11 (14.1)	< 0.001
Migration of pain to RLQ	98 (97.9)	50 (64.1)	< 0.001
Elevated temperature > 37.3°C	40 (39.6)	7 (9.0)	0.039
Urinal ketone bodies	51 (50.5)	15 (19.2)	< 0.001
Nausea and vomiting	80 (79.2)	52 (66.7)	0.059
Leucocytes* (× 10 ⁹ /L)	17.69 ± 5.99	10.22 ± 3.80	< 0.001
Leucocytosis (> 10,000 mm ³)	93 (92.1)	32 (41.0)	< 0.001
Neutrophilia (> 70%)	93 (92.1)	22 (28.2)	< 0.001
C-reactive protein* (mg/dL)	61.17 ± 74.8	12.11 ± 25.8	< 0.001
Alvarado score [†]	8 (6–9)	3 (2–5)	< 0.001

Data presented as *mean ± standard deviation or †median (interquartile range). p < 0.05 is statistically significant. RLQ: right lower quadrant

coli was the most frequent pathogen (68.7%). The rate of post-surgical complications was 0.02%. Based on the histopathological outcome of appendectomy specimens, the false-positive appendectomy rate was 7.3% (n = 8). CT was performed in three cases with equivocal results.

Table II presents the US findings in the study sample. In the appendicitis group (n = 101), the appendix was visualised in 66 (65.3%) appendectomies. In cases where the appendix was not visualised, at least one secondary sign was present in 32 (31.7%) cases. Completely visualised appendices with secondary signs were present in 52 (51.5%) cases. The secondary signs were hyperaemia, echogenic fat, appendicoliths, and free fluid or collection. The appendix visualisation rate in the non-appendicitis group (n = 78) was 15.4% (n = 12), and secondary signs were present in 6.4% (n = 5) cases. Based on the histopathological results, the ROC curve was used to evaluate the discriminative power of US examination, with an area under the curve (AUC) of 0.807 (95% confidence interval [CI] 0.642–0.972).

Based on the ROC analysis of laboratory results with continuous values and the appendix diameter measured in the US examination, the first step was to determine the thresholds to differentiate between patients with and without appendicitis. The results are presented in Fig. 2. Continuous variables were converted into categorical predictors, with the most discriminatory cut-off values as follows: 12.50 × 10⁹/L for WBC, 15 mg/dL for CRP and 7 mm for appendix diameter. The second step was to perform a logistic regression analysis on the US signs to determine the predictors of acute appendicitis. Blumberg's sign, free fluid or collection in the pouch of Douglas, hyperaemia, non-compressible appendix and diameter > 7 mm were found to be significant US predictive factors in paediatric acute appendicitis.

Table II. Ultrasonographic findings in the appendicitis and non-appendicitis groups.

Elements	No. (%)		p-value
	Appendicitis (n = 101)	Non-appendicitis (n = 78)	
Appendix diameter* (mm)	0.91 ± 0.24	0.62 ± 0.17	0.006
Appendix diameter > 7 mm	63 (62.3)	2 (2.6)	< 0.001
Non-compressible appendix	40 (39.6)	1 (1.3)	< 0.001
Hyper-/hypovascularisation	7 (6.9)	0	< 0.001
Adjacent fat inflammation	40 (39.6)	2 (2.6)	< 0.001
Appendicolith	3 (3.0)	0	< 0.001
Hyperaemia	55 (54.4)	1 (1.3)	< 0.001
Free fluid or collection	50 (49.5)	2 (2.6)	< 0.001
Mesenteric lymphadenopathy	0	34 (43.6)	< 0.001

*Data presented as mean ± standard deviation. p < 0.01 denotes statistical significance.

They were included in our US coefficient. An inflammatory coefficient was designed to capture the predictive capability of WBC, neutrophilia, CRP and ketone bodies. A mesenteric coefficient was included in the scoring system, based on the consideration that mesenteric lymphatic alterations can rule out acute appendicitis. A risk score (POPs) with the following formula was proposed:

$$POPs = \text{inflammatory coefficient} \times \text{US coefficient} \times \text{mesenteric coefficient}$$

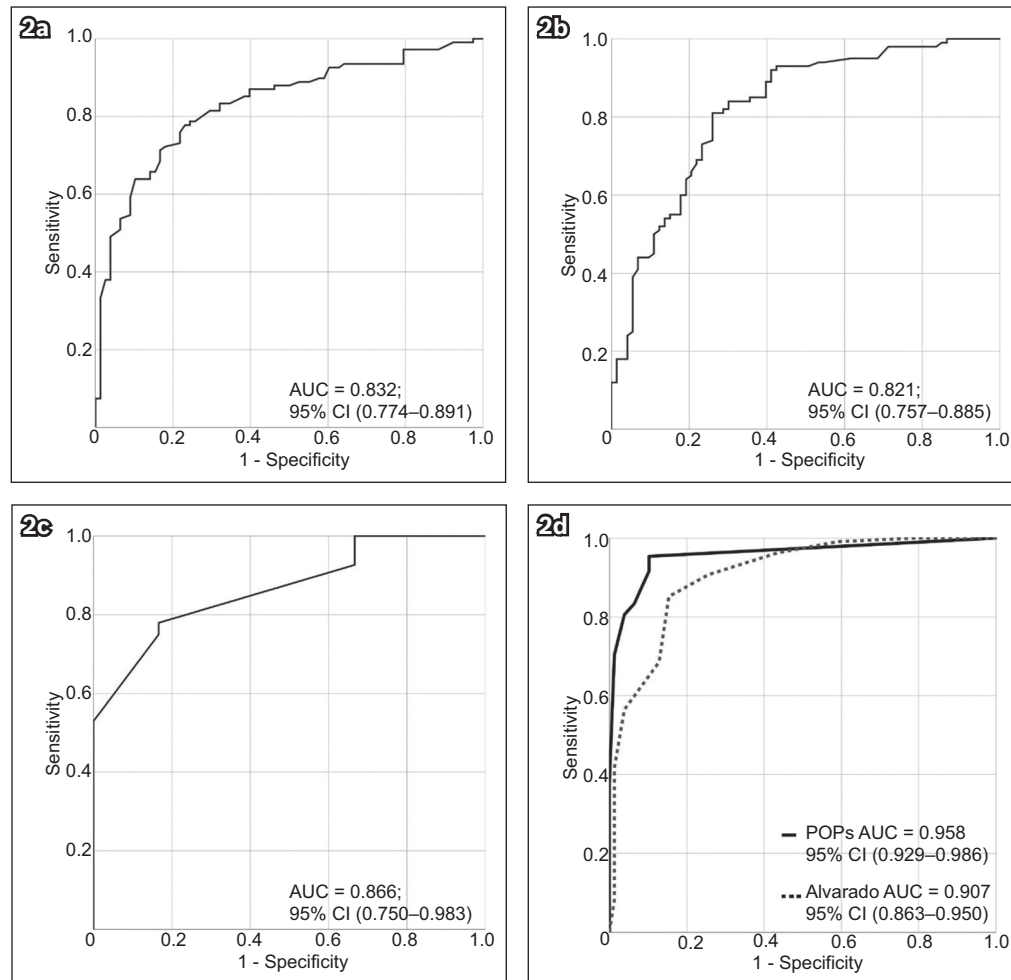


Fig. 2 Graphs show the receiver operating characteristic (ROC) analysis of the laboratory results for (a) leucocytes; (b) CRP levels; (c) appendiceal diameter; and (d) Alvarado score and POPs. AUC: area under the curve; CI: confidence interval; CRP: C-reactive protein

Table III summarises the elements of the POPs scoring system. Every element of the first two coefficients is worth 1 point, if present. The inflammatory coefficient totals up to 4 points and the US coefficient up to 5 points. The mesenteric coefficient is worth 1 point if lymphadenopathy is absent. The final score ranges from 0 to 20 points, and is obtained by multiplying the coefficients. The ROC for the proposed POPs is shown in Fig. 2d, with an AUC of 0.958 (95% CI 0.929–0.986).

The elements of the inflammatory and US coefficients add up to a maximum of 4 and 5 points, respectively. The final POPs score is the product of the three components, resulting in a possible range from 0 to 20 points. Mesenteric lymphadenopathy is decisive for a non-zero POPs score.

Two workable threshold values were identified, aimed at further risk stratification into three diagnostic zones for POPs: low-risk < 2.5 points; intermediate risk 2.5–7 points; and high-risk > 7 points. As observed in the study sample, the estimated safe low-risk interval had a 100% negative predictive value and the high-risk zone had a 100% positive predictive value for diagnosing child acute appendicitis (Fig. 3).

Fig. 2d shows that the AUC for the POPs scoring system was greater than that for the Alvarado score (0.958 [95% CI 0.929–0.986] vs. 0.907 [95% CI 0.863–0.950]), although the difference was not statistically significant ($p = 0.185$). However,

it is worth mentioning that if a similar risk stratification strategy was applied for the Alvarado score, it would falsely diagnose 10 (5.6%) cases with appendicitis (i.e. false positives) and miss 5 (2.8%) cases (i.e. false negatives).

DISCUSSION

Diagnosis of acute appendicitis in children is difficult owing to many factors, and the removal of a healthy appendix is associated with a greater risk of abdominal adhesions as compared to acute appendicitis.⁽¹⁵⁾ This is a hazard to be considered in contrast with an increasing rate of appendiceal perforation in delayed surgical interventions. Moreover, surgery should not be delayed in paediatric patients with a high suspicion of appendicitis owing to the high risk of perforation and further secondary complications.

Regardless of the surgical technique (laparotomy or laparoscopic approach), the scars acquired in childhood should also be considered when assessing the advantages and disadvantages of surgical intervention. These factors have been proven to have an impact in adulthood, as physical aesthetics might be important in the formation of personality.⁽¹⁶⁾

Disease prevalence in a patient population directly affects the positive predictive values of a diagnostic test, with low prevalence values leading to dramatic increases in false positives. In addition,

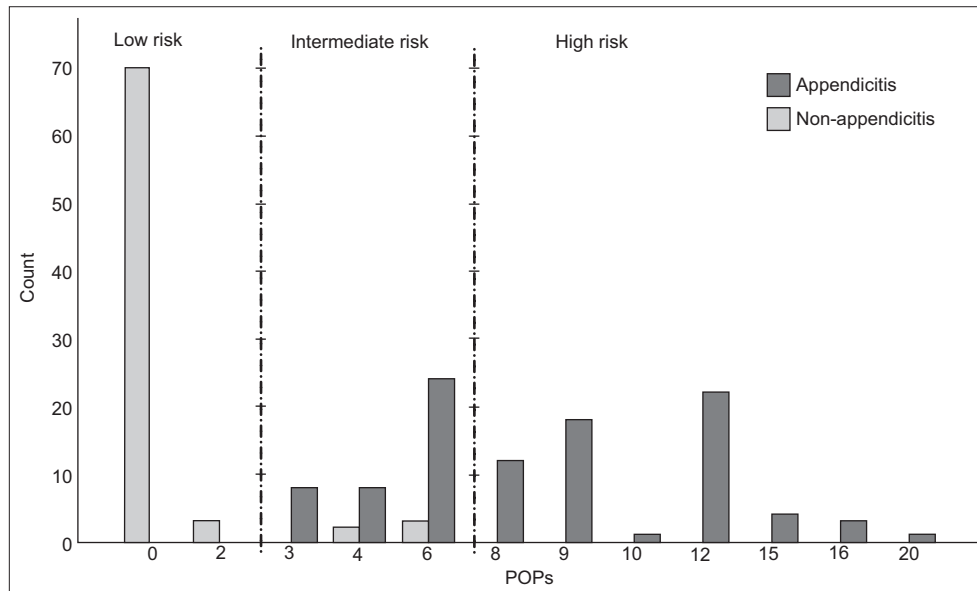


Fig. 3 Graph shows the negative and positive predictive value of the POPs scoring system.

Table III. Elements of the POPs scoring system.

POPs element	Point
Inflammatory coefficient	
Leucocytes > 12.50 × 10 ⁹ /L	1
Neutrophilia	1
CRP > 15 mg/dL	1
Ketone bodies	1
Ultrasound coefficient	
Blumberg's sign	1
Free fluid or collection	1
Hyperaemia	1
Non-compressible appendix	1
Appendix diameter > 7 mm	1
Mesenteric lymphadenopathy	
Present/absent	0/1

when the prevalence is low, infrequent cases of appendicitis may lead the radiologists to lose their diagnostic skills over time, thus worsening the performance of the US examination itself, despite US being widely accepted as a worthwhile diagnostic instrument.^(17,18) The visualisation rate of the appendix in our study was 43.6% (66 and 12 patients in the appendicitis and non-appendicitis groups, respectively), which was similar to the rate reported by Mittal et al,⁽¹⁹⁾ who found a lower rate of appendix visualisation in hospitals where US investigation is less often conducted (25%) as compared to hospitals where US is routinely employed in the diagnosis of appendicitis (56%). Trout et al also emphasised the high false negative and false positive rates in US examinations.⁽²⁰⁾ To improve diagnostic performance, the involvement of experienced personnel and/or additional training would be required.

US findings have been used to enhance the diagnosis methods of acute appendicitis in many studies, although a quantitative scoring system for suspected appendicitis in children is not yet available. Larson et al, in a study that evaluated the diagnostic

accuracy of US of the paediatric appendix among 1,357 examinations, reported a 96.8% accuracy of a five-category interpretive scheme.⁽²¹⁾ Although we did not classify US into categories in our study, we found a discriminative power of US by AUC (0.807 [95% CI 0.642–0.972]).

To the best of our knowledge, this is the first study to combine US and inflammatory results into a risk scoring system aimed at improving the diagnostic accuracy in emergency paediatric cases. In this study, only three cases had an appendicolith in the appendix; this factor was not found as a significant predictor for the diagnosis of appendicitis. However, in such a situation, calculating a score would obviously be redundant. On the other hand, larger samples might lead to further adjustments in the POPs formula.

We combined inflammatory markers and US findings with the aim of reducing the rate of false appendectomies in emergency cases. The proposed risk score proved to have better discriminatory capabilities than the existing Alvarado score. In a study of 1,235 patients, Fallon et al stratified the risk of appendicitis by US findings and found that US decreased the negative predictive value of appendicitis.⁽²²⁾ In our study, the US findings incorporated in POPs proved to be superior to the Alvarado score when stratification strategy was applied. Fig. 4 presents the scoring scheme that we proposed for paediatric acute appendicitis. When the initial US examination is equivocal, we suggest a clinical reassessment, followed by another US re-examination and surgical consultation.

A quantitative scoring for suspected appendicitis is only the first step towards improving the predictive values in the decision for surgical intervention and compliance with the GRADE recommendations for diagnostic tests and strategies.⁽²³⁾ The present study brought evidence of ameliorated accuracy in diagnosis and put forward the POPs-based scoring scheme for further investigation and testing in prospective studies.

This study was not without limitations. First, the different experience levels of the medical personnel conducting the

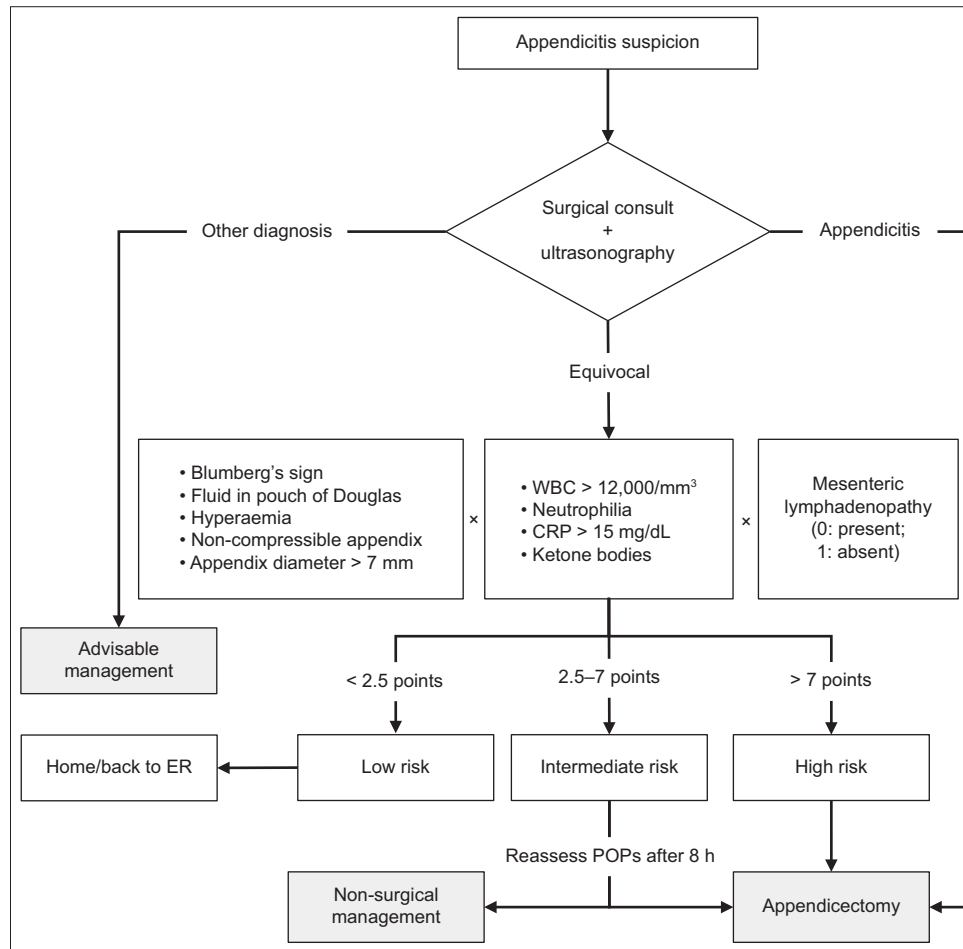


Fig. 4 Chart shows the proposed POPs scoring scheme for paediatric acute appendicitis. CRP: C-reactive protein; ER: emergency room; WBC: white blood cell

US examination may have led to wide margins for the rate of false negatives or false positives in the reported examinations. As our hospital does not have a standardised US protocol for appendicitis, the number of admitted patients who actually underwent US examination and could be included in this retrospective study was limited. Also, we used the same dataset for creating and validating the score performance. According to the TRIPOD statement, this model falls into Type 1a.⁽²⁴⁾ Another limitation is our approach of using only two thresholds for the two-way ROC instead of a three-way ROC. However, we believe this proposed practical solution to be workable and effective in real practice.

In conclusion, this study proposed a newly developed scoring scheme that integrates inflammatory predictors and US findings into a comprehensive score, allowing for effective risk stratification in paediatric cases of acute appendicitis, with a range of 0–20 risk points. US examination is minimally invasive and is thus a strongly recommended imaging procedure for paediatric investigations. This is on condition of an acceptable diagnostic accuracy, which makes the POPs-based scoring scheme a promising alternative to existing scores. Although further calibration would certainly be beneficial, the proposed scoring scheme is simple and easy to understand, remember and apply in the emergency room.

REFERENCES

1. Svensson JF, Patkova B, Almström M, Eaton S, Wester T. Outcome after introduction of laparoscopic appendectomy in children: a cohort study. *J Pediatr Surg* 2016; 51:449-53.
2. Zaidan H, Khalfan F, Ahmed H, Corbally MT. Positive and negative rates in children with acute appendicitis. *Bahrain Med Bull* 2018; 40:82-5.
3. Dokumcu Z, Toker Kurtmen B, Divarci E, et al. Retrospective multivariate analysis of data from children with suspected appendicitis: a new tool for diagnosis. *Emerg Med Int* 2018; 2018:4810730.
4. Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med* 1986; 15:557-64.
5. Samuel M. Pediatric appendicitis score. *J Pediatr Surg* 2002; 37:877-81.
6. Andersson M, Andersson RE. The appendicitis inflammatory response score: a tool for the diagnosis of acute appendicitis that outperforms the Alvarado score. *World J Surg* 2008; 32:1843-9.
7. Yap TL, Chen Y, Low WW, et al. A new 2-step risk-stratification clinical score for suspected appendicitis in children. *J Pediatr Surg* 2015; 50:2051-5.
8. Puylaert JB. Mesenteric adenitis and acute terminal ileitis: US evaluation using graded compression. *Radiology* 1986; 161:691-5.
9. Dibble EH, Swenson DW, Cartagena C, Baird GL, Herliczek TW. Effectiveness of a staged US and unenhanced MR imaging algorithm in the diagnosis of pediatric appendicitis. *Radiology* 2018; 286:1022-9.
10. Helbling R, Conficconi E, Wyttenbach M, et al. Acute nonspecific mesenteric lymphadenitis: more than "no need for surgery." *Biomed Res Int* 2017; 2017:9784565.
11. Shogilev DJ, Duus N, Odum SR, Shapiro NI. Diagnosing appendicitis: evidence-based review of the diagnostic approach in 2014. *West J Emerg Med* 2014; 15:859-71.
12. Chen CY, Zhao LL, Lin YR, Wu KH, Wu HP. Different urinalysis appearances in children with simple and perforated appendicitis. *Am J Emerg Med* 2013; 31:1560-3.
13. Inal M, Unal B, Bilgili YK. Better visualization of vermiform appendix with tissue harmonic imaging compared to conventional sonography. *Iran J Radiol* 2014; 11:e18114.

14. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; 44:837-45.
15. Memon ZA, Irfan S, Fatima K, Iqbal MS, Sami W. Acute appendicitis: diagnostic accuracy of Alvarado scoring system. *Asian J Surg* 2013; 36:144-9.
16. Mekeres F, Voita GF, Mekeres GM, Bodog FD. Psychosocial impact of scars in evaluation of aesthetic prejudice. *Rom J Leg Med* 2017; 25:435-8.
17. Khan U, Kitar M, Krichen I, et al. To determine validity of ultrasound in predicting acute appendicitis among children keeping histopathology as gold standard. *Ann Med Surg (Lond)* 2018; 38:22-7.
18. Krishnamoorthi R, Ramarajan N, Wang NE, et al. Effectiveness of a staged US and CT protocol for the diagnosis of pediatric appendicitis: reducing radiation exposure in the age of ALARA. *Radiology* 2011; 259:231-9.
19. Mittal MK, Dayan PS, Macias CG, et al. Performance of ultrasound in the diagnosis of appendicitis in children in a multicenter cohort. *Acad Emerg Med* 2013; 20:697-702.
20. Trout AT, Sanchez R, Ladino-Torres MF, Pai DR, Strouse PJ. A critical evaluation of US for the diagnosis of pediatric acute appendicitis in a real-life setting: how can we improve the diagnostic value of sonography? *Pediatr Radiol* 2012; 42:813-23.
21. 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; 204:849-56.
22. Fallon SC, Orth RC, Guillerman RP, et al. Development and validation of an ultrasound scoring system for children with suspected acute appendicitis. *Pediatr Radiol* 2015; 45:1945-52.
23. Schünemann HJ, Oxman AD, Brozek J, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008; 336:1106-10.
24. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; 162:W1-73.