Original Article

A New Score for the Diagnosis of Complicated Appendicitis in Children - Complicated Appendicitis Pediatric Score

Adelais K. Tzortzopoulou1,2, Mariza Tsolia2, Nikolaos Spyridis2, Panagiota Giamarelou4, Rodanti Sfakiotaki4, Alexander Passalides2, Nikolaos Zavras2

12nd Department of Pediatric Surgery, “P. & A. Kyriakou” Childrens’ Hospital, Athens, 2Medical School, University of Athens, Departments of 3Pathology and 4Radiology, Children’s Hospital “P. & A. Kyriakou”, Athens, Greece

Aim: This study aims to construct and validate a new score for diagnosis of complicated appendicitis in children, complicated appendicitis pediatric score (CoAPS), to guide residents’ clinical decision-making on choosing the correct patients for immediate surgery, reducing the emergency negative surgeries.

Methods: This prospective observational study enrolled two cohorts of patients 5–15 years old. Four hundred and seven consecutive patients were enrolled for the derivation cohort. Demographic data, clinical features, and histopathology data were collected. The outcome measure was the histological diagnosis of gangrenous appendicitis with or without perforation. The score was next validated in a separate cohort of 312 consecutive patients who were classified according to their risk of complicated appendicitis. The diagnostic performance of the score and the potential for the risk stratification to select patients for diagnostic imaging, emergency operative management, and reduce emergency negative operation rates were quantified.

Results: A positive “jumping up” test, vomiting, white blood cell >13.5 × 10^3/ml, lymphocytes <18%, and C-reactive protein >50 mg/dl were independent predictors for complicated appendicitis. The final prediction model exhibited an area under the curve of 0.890 (95% confidence interval: 0.859–0.922). The low-risk group demonstrated high sensitivity (90.4%) for complicated appendicitis, while scores 6 or more were very specific (95%) for the disorder. Describing the potential utility of the score, emergency ultrasound imaging would have been postponed in 14.5% of patients (P = 0.0016), and emergency negative explorations would have been cut by 87%.

Conclusion: The CoAPS score could guide residents in emergency management of children with complicated appendicitis reducing hospitalizations and urgent surgeries.

Keywords: Children, complicated appendicitis, diagnostic score, emergency surgery

Introduction

Classic surgical teaching holds that appendicitis is a progressive disease causing peritonitis through a process of inflammation, necrosis, and perforation of appendix layers. Recent epidemiological and clinical data strongly support the concept of two distinct pathological entities of appendicitis with different courses and prognoses.1,2 Simple appendicitis in children can be treated successfully with antibiotics in 92% of cases, while complex appendicitis proceeds rapidly to perforation and should be managed with

Access this article online

Quick Response Code:

Website: www.jiaps.com

DOI: 10.4103/ijiaps.jiaps_110_21

Address for correspondence: Dr. Nikolaos Zavras, Medical School, National and Kapodistrian University of Athens, Mikras Asias 75, 11527 Athens, Greece. E-mail: nzavras@med.uoa.gr

How to cite this article: Tzortzopoulou AK, Tsolia M, Spyridis N, Giamarelou P, Sfakiotaki R, Passalides A, et al. A new score for the diagnosis of complicated appendicitis in children – Complicated appendicitis pediatric score. J Indian Assoc Pediatr Surg 2022:27:410-8.
emergency surgery.[13] One can envision the information being helpful to an emergency physician or surgeon who is treating pediatric patients suspected of appendicitis and must decide to proceed either with antibiotics-first therapy or an appendectomy.

The critical factor of success of antibiotics-first therapy is the careful patient selection and the ability to exclude cases with a gangrenous appendix wall that is imminent to perforate or has already ruptured into the free peritoneal cavity. The current practice promotes imaging of the appendix as a key first step in the diagnostic workup of appendicitis in children.[4] Simple appendicitis is accurately diagnosed with ultrasound or CT scan imaging, but both of these modalities have limited sensitivity in ruling out cases of complex appendicitis.[5-7] This raises concern as the initial assessment with imaging of every patient suspected of appendicitis might put children at risk of nonoperative treatment failure or exposure to radiation. Magnetic resonance imaging can differentiate perforated from nonperforated appendicitis, but the costs, availability, and possible need for sedation prevent its use as the initial imaging modality in cases of acute pediatric appendicitis.[8,9]

Risk stratification of patients suspected of appendicitis with clinical scoring systems has been proved advantageous because it allows stratifying planning for imaging.[10] In pediatric patients, clinical scores may rule out a diagnosis of appendicitis minimizing the need for imaging. Unfortunately, they cannot predict disease severity with sufficient accuracy.[11] Scores use as diagnostic criteria several clinical symptoms and signs, such as abdominal tenderness, guarding, anorexia, nausea, and migration of pain. These may not be interpreted objectively, and interrater variability limits their reproducibility and clinical usefulness.[12] On the other hand, rebound tenderness is a reliable sign of peritonism, due to complex appendicitis and should be properly examined.[13] A judicious clinical approach from experienced examiners is required since rebound tenderness can be extremely painful in children and make them uncooperative and apprehensive.

In a previous paper, we have described a new clinical test, the “jumping up” test (J-up test) for pediatric patients with acute right lower guardant (RLQ) pain suspected for appendicitis, and introduced it as a surrogate for rebound tenderness. J-up test interprets face and body expressions due to mechanical stress imposed on the peritoneum when asking children to jump up and raise both their hands trying to reach a toy hanging down from the ceiling of the emergency room. J-up test is based on a simple technique of active distraction that is very acceptable by children, especially those who are “white robe” distressed. It might also be used in telemedicine as it can be recorded, transmitted, and evaluated by remote examiners.[14]

The primary aim of the current study was to construct a new clinical score system and stratify prospectively a cohort of patients with acute RLQ pain according to their risk of complicated appendicitis: either gangrenous or perforated. A clinical score with sufficient accuracy to select those pediatric patients, who require timely intervention, would have a beneficial effect on surgical caseload, postoperative complications, length of hospital stay, readmission rates, and total expenditure per patient. The second aim of our study was to validate prospectively our new score in a different cohort of patients, to match our junior residents’ ability to select complicated appendicitis patients at presentation with reasonable use of imaging, and facilitate better-informed decisions regarding emergency surgery or nonemergency management (NEM). The third aim of our study was to reduce the number of emergency surgeries without degrading the quality of care and use it in the future in the best way of selecting only the true-needed surgical management patients.

**Methods**

**Study design and outcome**

The present study was a prospective observational cohort study aimed at collecting data to develop a new clinical score system to stratify children with nontraumatic right lower quadrant (RLQ) pain according to their risk of complicated appendicitis. The differentiation between nonappendicitis and appendicitis was not the goal of this study. Given the emergency management of patients with complicated appendicitis, the aim was to develop a clinical score system that would promptly identify those patients who are at high risk of emergency surgery due to complicated appendicitis and give them a priority for operative management during the first 12 h after admission at the emergency department.

The study consisted of derivation and validation phases. The target population consisted of all consecutive patients between the ages of 5 and 15 years admitted at children’s emergency surgical department and complaining about acute RLQ pain of more than 2 h and <96 conservatively h duration. Exclusion criteria included patients with abdominal trauma, previous appendectomy or other abdominal surgery, and chronic medical or malignant conditions. Children with clinical or radiological suspicion of appendicitis who were conserva managed and children with dismissed follow-up were also excluded.

The Complicated Apendicitis Score (CoApPS) was developed based on the data from the derivation cohort
of 437 consecutive patients; 407 of them were eligible patients. The characteristics of subjects from derivation and validation cohorts including (1) demographic data (i.e., age, gender) (2) symptoms and signs (i.e., duration of symptoms, body temperature greater or lower than 37.5°C, RLQ tenderness, psoas sign, Rovsing sign, vomiting), (3) The jumping-up test, qualified by residents either as positive or negative, (4) imaging studies performed by staff radiologists, (5) operative notes, (6) histopathology reports. The white blood cell count (WBC), neutrophil percentage (NE), lymphocyte percentage (LY), platelet count (PLT), and C-reactive protein (CRP) data were tested on admission on venous samples were also collected. The discriminative and predictive ability of the developed score was analyzed in the derivation cohort and then validated in a different cohort of 323 consecutive patients, 312 of them were finally eligible at the same hospital. Derivation data were collected by one senior pediatric surgery resident, whereas data of the validation cohort were collected by on-duty junior residents with <1 year of experience in pediatric emergency surgery. The decision to operate was confirmed by a senior surgical staff member blinded to the results of the score. Imaging through ultrasound scanning was used selectively and at the discretion of the senior surgeon.

The interesting outcome was acute complicated appendicitis by histopathological diagnosis for operative patients. Complicated appendicitis is defined as gangrenous based on the presence of extensive necrotic tissue in the muscular layer of the appendix with or without signs of perforation. The diagnosis of simple (phlegmonous or suppurative) appendicitis entailed the presence of transmural and focal infiltrates or at least mucosal infiltrates of neutrophils without necrosis or perforation. For patients without appendicitis the alternative final diagnoses were categorized into complicated nonappendicitis diseases, suitable for surgery, (i.e., Meckel diverticulitis), and uncomplicated nonappendicitis diseases (suitable for conservative management). For patients with a diagnosis other than appendicitis (“nonoperated nonappendicitis” group of patients), a telephone interview with their caregivers was made to confirm the final diagnosis 6 weeks after visiting. The data collection and research protocol were approved by the ethics board of the hospital.

Statistical analysis
Excel software was used for data entry. Four hundred and seven patients of the derivation cohort were categorized into two groups. The first group included patients with acute RLQ pain and histological diagnosis of complicated appendicitis (gangrenous and/or no perforation), and the second group included patients with RLQ pain of any other final diagnosis. Univariate analysis was utilized to determine the effect of potential factors on complicated appendicitis. The normal distribution of data was evaluated with the Shapiro–Wilk test. Numerical values without normal distribution were presented as medians and interquartile ranges. Categorical variables were presented as numbers and percentages. Two-tailed Fisher’s exact test was used in a comparison of categorical data. Numerical values, if not normally distributed, were compared using the Mann-Whitney U test between groups. If numerical data were not normally distributed and did not have homogeneity of variances according to Levene’s test, then one-way ANOVA and Brown–Forsythe’s robust test for equality of means was used.

Construction and validation of the score
To enhance clinical applicability continuous variables were converted into categorical variables using ROC curve analysis to obtain the best cut-off points to differentiate patients with complicated appendicitis from all other patients with acute RLQ pain. The highest diagnostic odds ratio (DOR, Likelihood ratio + ve/ Likelihood Ratio –ve) was chosen as the optimal cut-off value for each statistically significant continuous variable.

A binary forward (likelihood ratio) stepwise logistic regression analysis included the categorized significant factors, and their corresponding weightage (β-coefficients) were identified. The discriminative capacity of the prediction model was expressed as an area under the ROC curve (AUC) with corresponding 95% confidence intervals (95% CI). The model was transformed in a clinically applicable scoring system obtaining points from regression β-coefficients of the significant predictors rounded to the nearest integer. The diagnostic performance of the score was validated by a separate cohort of 312 patients and analyzed in 2 × 2 contingency tables and receiver operator characteristic AUC. As the aim of our study was to correctly identify as many patients with complicated appendicitis a possible while limiting the number of false negatives, we classified patients into three risk categories according to ROC curve analysis: High risk (specificity over 85%, sensitivity at least 20%), intermediate-risk (sensitivity almost equal to specificity) and low risk (sensitivity over 90%, specificity at least 30%). The clinical utility of the score was finally estimated by its potential outcome on complicated appendicitis rates, imaging rates, emergency operations rates, and negative explorations rates as if it has been applied in daily clinical practice.
Statistical package SPSS 25 windows 64 was used for statistical assessment and drawing ROC curves.

**RESULTS**

The entire number of patients who met the inclusion criteria during the time frame of the study was 719. We included 407 patients in the derivation cohort and 312 patients in the validation cohort. Data on patients' management and the final diagnosis of the two cohorts are presented in Table 1. There was a statistically significant difference in explorations rate between patients in derivation and validation cohorts (51.59% vs. 66.98% respectively, \( P = 0.0001 \)). There were no differences in complicated appendicitis rates, simple appendicitis rates, negative appendectomies, and explorations rates in patients with non-complicated appendicitis diagnosis.

The clinical characteristics of derivation and validation cohorts are listed in Table 2. No significant differences in age, gender, duration of RLQ pain, fever, vomiting, and psoas sign existed between cohorts. Patients in the validation cohort were reported as being more positive to Rovsing’s sign more frequently (12.04% vs. 7.05%, \( P = 0.032 \)). There were also significant differences between derivation and validation cohorts in NE (74.30% vs. 77.40%, \( P = 0.019 \)), LY (17.50% vs. 15.30%, \( P = 0.022 \)), and ultrasound imaging (28.39% vs. 34.94%, \( P = 0.0052 \)).

In derivation cohort patients with complicated appendicitis were mostly males (63.64% vs. 49.30%, \( P = 0.0084 \)), had fever over 37.5°C (45.45% vs. 23.78%, \( P < 0.0001 \)), reported more frequently vomiting (64.46% vs. 56.75%, \( P = 0.035 \)) while patients in the validation set exhibited Rovsing’s sign more frequently (12.04% vs. 7.05%, \( P = 0.032 \)). There were also significant differences between age, gender, duration of RLQ pain, fever, vomiting, and psoas sign existed between cohorts. Patients in the validation cohort were reported as being more positive to Rovsing’s sign more frequently (12.04% vs. 7.05%, \( P = 0.032 \)). There were also significant differences between derivation and validation cohorts in NE (74.30% vs. 77.40%, \( P = 0.019 \)), LY (17.50% vs. 15.30%, \( P = 0.022 \)), and ultrasound imaging (28.39% vs. 34.94%, \( P = 0.0052 \)).

In derivation cohort patients with complicated appendicitis were mostly males (63.64% vs. 49.30%, \( P = 0.0084 \)), had fever over 37.5°C (45.45% vs. 23.78%, \( P < 0.0001 \)), reported more frequently vomiting (64.46% vs. 56.75%, \( P = 0.035 \)) while patients in the validation set exhibited Rovsing’s sign more frequently (12.04% vs. 7.05%, \( P = 0.032 \)). There were also significant differences between derivation and validation cohorts in NE (74.30% vs. 77.40%, \( P = 0.019 \)), LY (17.50% vs. 15.30%, \( P = 0.022 \)), and ultrasound imaging (28.39% vs. 34.94%, \( P = 0.0052 \)).

The score variables and the respective score points included in the analysis were: Male gender, duration of pain <20 h, temperature >37.5°C, Jumping-up test + ve, vomiting, psoas sign, WBC >13.5 × 10^3 µL, NE >83%, LY <18% and CRP >50 mg/dL. Predictors finally included in the analysis were: Jumping up test + ve, vomiting, WBC >13.5 × 10^3 µL, LY <18%, and CRP >50 mg/dL [Table 3]. Diagnosis of collinearity for the above variables was performed, and the variance inflation factors were 1.257, 1.142, 1.446, 1.568, and 1.048, respectively, suggesting that there was no multiple collinearity relationship.

ROC curve analysis of the prediction model resulted in an AUC of 0.890 (95% CI: 0.859–0.922, \( P < 0.05 \)). The fitness of the model was evaluated using the Hosmer–Lemeshow goodness-fit test (\( P > 0.05 \)). The value of \( x^2 \) was 0.457, suggesting that the model fitted well.

The score variables and the respective score points are presented in Table 4. Hence, 3 points were given when the jumping-up test was positive, 1 point in vomiting, 1 point when WBC was >13.5 × 10^3/L, 1 point when LY was lower than 18% and CRP was >50 mg/dL. The total score of complicated appendicitis pediatric score was 7.

**Diagnostic performance of CoApPS**

The constructed score was validated by a separate cohort of 312 eligible patients. The AUC for the detection of all cases of complicated appendicitis in the validation set was 0.844 (95% CI: 0.798–0.880, \( P < 0.05 \)), demonstrated good discriminative capacity. The diagnostic performance of the score in the validation cohort is presented in Table 5. The cut-off score of 6, for differentiating the high–risk group of patients with complicated appendicitis yielded a specificity of 94.9% and a positive predictive value of

**Table 1: Patients’ management and final diagnosis**

|                      | Derivation cohort | Validation cohort | \( P \)  |
|----------------------|-------------------|------------------|---------|
| Sample size          | 407               | 312              |         |
| Number of explorations, \( n \) (%) | 210 (51.59)       | 209 (66.98)      | 0.0001^1 |
| Complicated appendicitis, \( n \) (%)* | 125 (30.71)       | 116 (37.17)      | 0.0794^4 |
| Simple appendicitis, \( n \) ** | 71 (17.44)        | 70 (22.43)       | 0.0882^3 |
| Negative appendectomies (percentage of explorations) | 9 (4.28)          | 15 (7.18)        | 0.2153^3 |
| Operated on for other diagnosis (percentage of explorations)^3 | 7 (3.33)          | 8 (3.82)         | 0.7998^3 |
| Nonappendicitis, nonoperative management, \( n \) | 197 (48.40)       | 103 (33.01)      | 0.0001^1 |

*Gangrenous with or without perforation, **Phlegmonous, suppurative, operative management, ^i.e., Meckel’s diverticulitis, ovarian cyst. In all cases an appendectomy was performed, ^Two-tailed Fisher’s exact test
Tzortzopoulou, et al.: CoApPS: Diagnosing complicated appendicitis in children

Based on the chosen cut-off values in the ROC analysis, patients were stratified into three risk categories corresponding to posttest probability of complicated appendicitis: High risk (≥6 points), intermediate-risk (4-5 points), low risk (0-3 points) [Figures 1, 2 and Table 6]. The score designated correctly 56 of 116 (48.2%) patients with complicated appendicitis as high-risk patients. Into the high-risk group were also assigned 8 patients with simple appendicitis and one patient operated for a complicated adnexal cyst. Seventeen of 116 patients (14.65%) were classified into the low-risk category. Of the 103 not operated patients, 92 (89.32%) were correctly assigned into the low-risk group. There were 8 (53.33%) of the 15 negative explorations performed in low-risk patients giving a negative exploration rate of 13.11% (8 out of 61) within this group. Histopathology results of risk categories are presented in Figure 3.

Potential clinical utility of the CoApPS score
Regarding the potential clinical utility of the score, if high-risk group patients were admitted as candidates for emergency surgery during the next 12 h after admission at the emergency department and if emergency imaging with ultrasound scanning was performed only in the intermediate-risk group of patients, emergency operations would have been reduced by 36% while 82% of complicated appendicitis. 52% of simple appendicitis patients would have been managed with emergency surgery. In addition, emergency ultrasound

| Table 2: Clinical and imaging characteristics of patients |
|----------------------------------------------------------|
| **Predictors**                                            | **Derivation cohort (n=407)** | **Validation cohort (n=312)** | **P** |
| Median age (IQR)                                         | 10 (8-12)                     | 11 (9-13)                     | 0.050† |
| Males, n (%)                                             | 216 (53.07)                   | 167 (53.52)                   | 0.820† |
| Duration of pain (hours, median, IQR)                    | 18 (11-48)                    | 24 (11-30)                    | 0.516† |
| Temperature >37.5°C (%)                                  | 123 (30.22)                   | 111 (35.57)                   | 0.108† |
| Vomitus, n (%)                                           | 149 (36.60)                   | 127 (40.70)                   | 0.183† |
| Jumping up test positive, n (%)                          | 231 (56.75)                   | 201 (64.42)                   | 0.035† |
| Psoas sign, n (%)                                        | 90 (22.11)                    | 70 (22.44)                    | 0.094† |
| Rovsing sign, n (%)                                      | 49 (12.04)                    | 22 (7.05)                     | 0.032† |
| WBC count, median (IQR) ×10³/μL                          | 11.9 (8.6-15.8)               | 12.6 (9-16.7)                 | 0.440† |
| NE %, median (IQR)                                       | 74.30 (59.3-82.5)             | 77.40 (64.45-84.90)           | 0.019† |
| LY%, median (IQR)                                        | 17.50 (10.05-31.45)           | 15.30 (8.75-25.80)            | 0.022† |
| CRP, median (IQR) mg/dL                                  | 5 (1-21)                      | 6 (1-22.25)                   | 0.976† |
| PLT, median, (IQR) (×10³ μL)                             | 281 (240-323)                 | 276 (1-22.25)                 | 0.256† |
| Imaging (ultrasonography), n (%)                         | 118 (28.99)                   | 109 (34.94)                   | 0.0052† |

†Two-tailed Fisher’s exact test, ‡One way ANOVA and Brown-Forsyth test. IQR: Interquartile range, WBC: White blood cell count, LY: Lymphocyte, CRP: C-reactive protein, CI: Confidence interval, SE: Standard error

| Table 3: Binary logistic regression analysis of significant predictors for complicated appendicitis (derivation cohort, n=407) |
|----------------------------------------------------------------------------------|
| **Predictors** | **B** | **SE** | **Wald** | **df** | **Significance** | **Exp(B)** | **Lower** | **Upper** |
| Jumping up test positive            | 3.076 | 0.559 | 82.09    | 1      | 0.000        | 21.671    | 8.165    | 57.521    |
| Vomiting                            | 1.323 | 0.304 | 18.906   | 1      | 0.000        | 3.753     | 2.068    | 6.813     |
| WBC count >13.5 (×10³ μL)           | 0.974 | 0.315 | 9.557    | 1      | 0.002        | 2.648     | 1.428    | 4.910     |
| LY <18%                             | 1.102 | 0.350 | 9.919    | 1      | 0.002        | 3.012     | 1.516    | 5.981     |
| CRP >50 mg/dL                       | 0.989 | 0.430 | 5.289    | 1      | 0.021        | 2.687     | 1.157    | 6.241     |
| Constant                            | −5.067| 0.559 | 82.090   | 1      | 0.000        | 0.006     |          |           |

*Diagnosis by histopathology: Gangrenous with or without perforation. WBC: White blood cell count, LY: Lymphocyte, CRP: C-reactive protein, CI: Confidence interval, SE: Standard error

| Table 4: The complicated appendicitis pediatric score |
|-----------------------------------------------------|
| **Predictors** | **Score** |
| Jumping up test positive                       | 3         |
| Vomitus                                        | 1         |
| WBC count >13.5 (×10³ μL)                      | 1         |
| LY (%) lower than 18                           | 1         |
| CRP >50 mg/dL                                  | 1         |
| Maximum score                                  | 7         |

WBC: White blood cells, LY: Lymphocytes, CRP: C-reactive protein

84.73% (likelihood of having complicated appendicitis in the high-risk group).

Based on the chosen cut-off values in the ROC analysis, patients were stratified into three risk categories corresponding to posttest probability of complicated appendicitis: High risk (≥6 points), intermediate-risk (4-5 points), low risk (0-3 points) [Figures 1, 2 and Table 6]. The score designated correctly 56 of 116 (48.2%) patients with complicated appendicitis as high-risk patients. Into the high-risk group were also assigned 8 patients with simple appendicitis and one patient operated for a complicated adnexal cyst. Seventeen of 116 patients with complicated appendicitis (14.65%) were classified into the low-risk category. Of the 103 not operated patients, 92 (89.32%) were correctly assigned into the low-risk group. There were 8 (53.33%) of the 15 negative explorations performed in low-risk patients giving a negative exploration rate of 13.11% (8 out of 61) within this group. Histopathology results of risk categories are presented in Figure 3.

Potential clinical utility of the CoApPS score
Regarding the potential clinical utility of the score, if high-risk group patients were admitted as candidates for emergency surgery during the next 12 h after admission at the emergency department and if emergency imaging with ultrasound scanning was performed only in the intermediate-risk group of patients, emergency operations would have been reduced by 36% while 82% of complicated appendicitis. 52% of simple appendicitis patients would have been managed with emergency surgery. In addition, emergency ultrasound
imaging would have been postponed in 14.5% of patients (P = 0.0016), emergency negative explorations would have been cut by 87%, and emergency negative exploration rates reduced from 7% to 1.5% (P = 0.0203). In addition, 18% of complicated appendicitis and 48% of simple appendicitis patients would have been considered for NEM, either operative or antibiotics-first therapy, during the next 12–24 h from admission at an emergency department.

**Discussion**

Delaying diagnosis of complicated appendicitis when hospital systems are reaching capacity is associated with higher complication rates, length of hospital stay, and heavier socioeconomic burden. Many factors, including increased caseload, diagnostic uncertainty, surgeon, and imaging availability, constitute to delay in surgery. Therefore, timely identification and emergent operative treatment of patients with gangrenous or perforated appendicitis would be an important issue to all pediatric caregivers.

Clinical score systems currently in practice have the potential to correctly stratify children at low risk of acute appendicitis and who could be considered for early discharge, but they cannot predict which children with appendicitis should proceed directly to surgery.

As a result, routine preoperative ultrasound and/or low dose computed tomography (CT) scan imaging is recommended in all intermediate or high-risk patients to secure a correct diagnosis. These patients may benefit from this strategy but only to rule in and not rule out complicated appendicitis. Ultrasound scan imaging demonstrates a high specificity but lower sensitivity for complicated appendicitis, and it has been reported that 56% of perforated appendicitis did not exhibit typical imaging findings. In addition, conventional simulated CT scan imaging failed to detect 79.5% of gangrenous non-perforated appendicitis as these cases were interpreted as uncomplicated appendicitis. Therefore if conservative treatment is chosen for these complicated cases, further imaging would be performed to rule in complicated appendicitis.

**Table 5:** Performance of the complicated appendicitis pediatric score (validation cohort, *n* = 312)

| Score | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) |
|-------|----------------|----------------|---------|---------|
| 0     | 100            | 0.00           | 37.18   | -       |
| 1     | 99.1           | 19.4           | 42.12   | 97.3    |
| 2     | 96.5           | 31.1           | 45.32   | 93.8    |
| 3     | 90.4           | 44.4           | 49.04   | 88.7    |
| 4     | 85.2           | 68.9           | 61.85   | 88.7    |
| 5     | 73.9           | 81.6           | 70.39   | 84.1    |
| 6     | 47.8           | 94.9           | 84.73   | 75.4    |
| 7     | 8.7            | 99.9           | 98.10   | 64.9    |

PPV: Positive predictive value, NPV: Negative predictive value

**Table 6:** Risk stratification according to the complicated appendicitis pediatric score for complicated appendicitis in children (validation cohort)

| Score   | Patient with RLQ pain (*n* = 312), n (%) | Nonoperated nonappendicitis (*n* = 103), n (%) | Operated complicated appendicitis (*n* = 116), n (%) | Operated simple appendicitis (*n* = 70), n (%) | Negative explorations (*n* = 15), n (%) | Operated for other diagnosis (*n* = 8), n (%) |
|---------|------------------------------------------|-----------------------------------------------|-----------------------------------------------------|-------------------------------------------|------------------------------------------|-------------------------------------------|
| High risk | 6-7                                      | 65 (20.83)                                    | 0 (0)                                                | 56 (48.27)                                | 8 (11.42)                                | 0 (12.5)                                  |
| Intermediate risk | 4-5                                      | 94 (30.13)                                    | 11 (10.68)                                           | 43 (37.06)                                | 31 (44.28)                               | 7 (46.67)                                 |
| Low risk | 0-3                                      | 153 (49.03)                                   | 92 (89.32)                                           | 61 (29.18)                                | 17 (14.65)                               | 31 (44.28)                                |

*Operated versus nonoperated patients. Fisher’s exact test. High versus intermediate risk: *P* < 0.0001, high versus low risk: *P* = 0.0001, intermediate versus low risk: *P* = 0.0815. *Operated with complicated appendicitis versus all other operated patients. Fisher’s exact test. High versus intermediate risk: *P* < 0.0001, high versus low risk: *P* = 0.0001, intermediate versus low risk: *P* = 0.006. *Negative versus nonnegative explorations. Fisher’s exact test. High versus intermediate risk: *P* = 0.0181, high versus low risk: *P* = 0.0023, intermediate versus low risk: *P* = 0.4148. RLQ: Right lower guardant
cases, it will be a higher rate of treatment failure and severe complications.

Pediatric patients with acute RLQ pain constitute a difficult puzzle of differential diagnosis since an emergency physician should answer promptly in two questions. First which patient with appendicitis must proceed directly to surgery and second which patient also with appendicitis might be better treated with antibiotics-first management.

Gorter et al. validated retrospectively a clinical score model in 65 children who had undergone surgery, intending to differentiate simple phlegmonous appendicitis from complicated cases, defined histologically as gangrenous with or without perforation. Ultrasound examination was performed in all patients. Duration of pain, temperature, diffuse guarding, CRP levels over 38 mg/L, and ultrasound signs of complex appendicitis were predictive factors of their score model.

The authors reported for their score a sensitivity of 90%, a specificity of 91%, a positive predictive value (PPV) of 64%, and a negative predictive value (NPV) of 98%. The authors concluded that their model could be used to exclude complicated appendicitis in patients hospitalized for suspected appendicitis if their score is 4. Ultrasound imaging was performed in all patients suspected of appendicitis. A limitation factor of this study is that the target population included only operated patients and not the entire population presented at the emergency department complaining about acute RLQ pain. The small number of patients should also be considered in the interpretation of the results.

Two recent studies have examined the diagnostic accuracy of appendicitis inflammatory response (AIR) and modified Heidelberg score in the diagnosis of perforated from nonperforated appendicitis in children. These studies defined as nonperforated the histologically phlegmonous together with gangrenous appendicitis cases. In a prospective study of AIR score in 184 children surgically treated with appendicectomy, the patients were allocated into two groups. One group included surgically treated patients with nonperforated appendicitis (histologically defined as phlegmonous or gangrenous appendicitis without perforation) and the second group included operated cases with perforation. The authors reported that 86.8% of perforated appendicitis were correctly placed in the high-risk group by an AIR score equal to or >9 while 13.2% of perforated appendicitis were incorrectly allocated in the intermediate-risk group. At the same time, 27.4% of nonperforated cases (simple phlegmonous or gangrenous without perforation) were allocated in the high-risk group and 67.1% in the intermediate risk group. Therefore the sensitivity of the AIR score for perforated appendicitis was 89.5%, the specificity was 71.9%, but the PPV was only 45.3% while the NPV was 96.3%. The authors concluded that the AIR score could differentiate perforated from nonperforated (simple phlegmonous or gangrenous) appendicitis in pediatric patients with a high level of accuracy. It should be noted that the AIR score uses as a predictive factor rebound tenderness/abdominal guarding graded in three levels of severity, light, medium, and strong. This could be interpreted rather subjectively by different examiners. AIR score also does not distinguish patients with phlegmonous
appendicitis from those with necrosis of the appendiceal wall (gangrenous) and therefore is not appropriate to be used in the triage selection of patients for antibiotics-first management.

In a retrospective analysis of a modified Heidelberg score (mHAS) in 423 children hospitalized at the pediatric surgical department, the patients were also allocated in two groups, one with histologically proven perforated appendicitis and another with nonperforated cases (simple phlegmonous or gangrenous). The mHAS correctly identified 97.1% of perforated appendicitis patients but only 17.9% of nonperforated appendicitis (simple phlegmonous or gangrenous) were correctly identified as such. The authors reported a high negative predictive value of 93.3% but a low positive predictive value of 34% for their score. They concluded that classical clinical signs may be less important than previously assumed in the diagnosis of pediatric appendicitis. They also stressed the importance of ultrasound as an essential component of the score and at the same time, remarked that the US can potentially delay surgical treatment. This was a retrospective study, and the target population included only hospitalized patients already diagnosed as having acute appendicitis. In our opinion, the modified Heidelberg score does not distinguish between simple phlegmonous appendicitis and gangrenous without perforation cases and therefore cannot be used in the decision making process for emergency surgery or antibiotics-first management of patients with acute RLQ pain.

In our prospective observational study, we developed a clinical score, the jumping up score, to quantify a patient’s risk of complicated appendicitis (gangrenous with or not perforation) when presenting with acute RLQ pain at the emergency department. Our model used as a predictive factor of peritonism a new clinical test—the Jumping up test (J-up) that can be interpreted objectively with the use of Bieri’s face pain scale. In our study, we defined complicated appendicitis in all cases histologically proven not only as perforated but also as gangrenous without perforation. This was done with the scope to identify those cases that were not suitable candidates for antibiotics-first management since perforation of the appendix wall is imminent in cases with a necrotic appendix wall, and antibiotics presumably could not prevent peritonitis. The target population of our study was all patients presented at the emergency surgical pediatric ward with acute RLQ pain and not only surgically treated patients with appendicectomy, as that was reported in the aforementioned studies with AIR and modified HAS scores. The aim of our study was not only to identify as many as possible of complicated cases (gangrenous or perforated) and treat them with emergency surgery, but at the same time, we intended to select as many as possible of noncomplicated cases (simple phlegmonous appendicitis without necrosis of appendix wall) with minimal use of ultrasound imaging. We validated the diagnostic accuracy of our new score and evaluated its potential benefit in the decision-making process concerning emergency admissions, diagnostic imaging, emergency operations, and negative emergency explorations.

The CoApP score less than 4 correctly classified into low-risk group 89% of overall RLQ patients with any other than complicated appendicitis diagnosis maintaining thus a high sensitivity (90.5%) for cases of complicated appendicitis. A score of 6 or greater classified correctly 48% of overall patients with complex appendicitis into the high-risk group, the score demonstrating a high-risk cut-off with excellent specificity of 94.9% for gangrenous and/or perforated appendicitis. The positive predictive value was 84.73%, and the negative predictive value was 75.4%. The few patients with high-risk scores and who did not have complicated appendicitis had conditions that otherwise necessitated hospital admission and operative management. The score, therefore, identifies patients at higher risk of morbidity who are likely to benefit from emergency operational management, defined as surgery during the next 12 h after admission to the emergency department. Diagnostic imaging in this group of patients may not be needed.

The CoApP score designated 37% of complicated appendicitis patients into the intermediate-risk group. We have estimated that if we hospitalize the subgroup of intermediate-risk patients with positive ultrasound findings to proceed with surgery during the next 12 h of their presentation at the emergency department, then 82% of overall complicated appendicitis patients would have been targeted correctly by our score. At the same time, emergency ultrasound imaging would have been performed only in one-third of the total number of patients. Almost half (47.8%) of simple appendicitis patients would also be considered for intravenous antibiotics-first management. Emergency negative operations, especially during night times, would have been avoided in 86.7% of total negative explorations.

**Conclusion**

Our observational data provide strong evidence that stratifying RLQ pain patients with CoApP scores at the presentation at the emergency department can guide decision-making to improve patient triage for emergency
surgery or antibiotics-first management with reasonable imaging utilization. Our new score could be proved advantageous since it favors a timely intervention with potentially lower complications, length of stay, cost, and lost days from school. Prompt and correct diagnosis of complicated appendicitis will allow us to make the best decisions regarding safer patient care, balanced with fair management of hospital imaging resources.

**Financial support and sponsorship**

Nil.

**Conflicts of interest**

There are no conflicts of interest.

**REFERENCES**

1. Andersson RE. The natural history and traditional management of appendicitis revisited: Spontaneous resolution and predominance of prehospital perforations imply that a correct diagnosis is more important than an early diagnosis. World J Surg 2007;31:86-92.

2. Livingston EH, Woodward WA, Sarosi GA, Haley RW. Disconnect between incidence of nonperforated and perforated appendicitis: Implications for pathophysiology and management. Ann Surg 2007;245:886-92.

3. Maita S, Andersson B, Svenson JF, Wester T. Non operative treatment for non perforated appendicitis in children: A systematic review and meta-analysis. Pediatr Surg Int 2020;36:261-9.

4. Bolmers MD, van Rossem CC, Gorter RR, Bemelman WA, van Geloven AA, Heij HA, et al. Imaging in pediatric appendicitis is key to a low normal appendix percentage: A national audit on the outcome of appendectomy for appendicitis in children. Pediatr Surg Int 2018;34:543-51.

5. Tseng P, Berdahl C, Kearl VL, Behar S, Cooper J, Dillbaum R, et al. Does right lower quadrant ultrasound accurately identify perforation in pediatric acute appendicitis? J Emerg Med 2016;50:638-42.

6. Riedesel EL, Weber BC, Shore MW, Cartmill RS, Ostlie DJ, Leys CM, et al. Diagnostic performance of standardized protocol for detecting perforation in pediatric appendicitis. Pediatr Radiol 2012;42:17-34.

7. Kim HY, Park HJ, Lee HJ, Lee SS, Jeon JJ, Lee KH. Systemic review and meta-analysis of CT features for differentiating complicated appendicitis. Radiology 2018;287:104-15.

8. Moore MM, Kulyalat AN, Hollenbeak CS, Engbrecht BW, Dillman JR, Methratta ST. Magnetic resonance imaging in pediatric appendicitis: A systemic review. Pediatr Radiol 2016;46:928-39.

9. James K, Duffy P, Kavanagh RG, Carey BW, Power S, Ryan D, et al. Fast acquisition abdominal MRI study for the investigation of suspected acute appendicitis in paediatric patients. Insights Imaging 2020;11:78.

10. Koberlein GC, Trout AT, Rigsby CK, Iyer RS, Alazraki AL, Anupindi SA, et al. ACR appropriateness criteria suspected appendicitis, child. J Am Coll Radiol 2019;16:S252-63.

11. van Amstel P, Gorter RR, Vander Lee JH, Cense HA, Bakx R, Heij HA. Ruling out appendicitis in children: Can we use clinical prediction rules? J Gastrointest Surg 2019;23:2027-48.

12. Kharbanda AB, Stevenson MD, Macias CG, Sinclair K, Dudley NC, Bennett J, et al. Interrater reliability of clinical findings in children with possible appendicitis. Pediatrics 2012;129:695-700.

13. Chang PH, Dai K, Yang Z, Wong KK. Validity of Alvarado score in predicting disease severity and postoperative complications in pediatric appendicitis. World J Pediatr Surg 2019;2:e000003.

14. Tzortzopoulou AK, Giamarelo P, Tsolia M, Spyridis N, Vakaki M, Passalides A, et al. The jumping up (J-Up) test: Making the diagnosis of acute appendicitis easier in children. Glob Pediatr Health 2019;6;2337794X19884824.

15. Kvasnovsky CL, Shi Y, Rich BS, Glick RD, Soffer SZ, Lipskar AM, et al. Limiting hospital resources for acute appendicitis in children: Lessons learned from the US epicenter of the Covid-19 pandemic. J Pediatr Surg. 2021;56:900-4. Doi: 10.1016/j.jpedsurg.2020.06.024.

16. Gerall CD, DeFazio JR, Kahan AM, Fan W, Fallon EM, Middleworth W, et al. Delayed presentation and sub-optimal outcomes of pediatric patients with acute appendicitis during the COVID-19 pandemic. J Pediatr Surg 2021;56:905-10. Doi: 10.1016/j.jpedsurg.2020.10.008. Epub 2020 Oct 19.

17. Papandria D, Goldstein SD, Rhee D, Salazar JH, Arlikar J, Gorgy A, et al. Risk of perforation increases with delay in recognition and surgery for acute appendicitis. J Surg Res 2013;184:723-9.

18. Carpenter JL, Orth RC, Zhang W, Lopez ME, Mangona KL, Guillerman RP. Diagnostic performance of US for differentiating perforated from non perforated pediatric appendicitis. A prospective cohort study. Radiology 2017;262:835-41.

19. Elbanna KY, Mohammed MF, Chalal T, Khosa F, Ali IT, Berger FH, et al. Dual energy CT in differentiating non perforated gangrenous appendicitis from uncomplicated appendicitis. Am J Roentgenol 2018;211:776-82.

20. Gorter RR, van den Boom AL, Heij HA, Knoopkens CM, Hulscher CC, Tenhagen M, et al. A scoring system to predict the severity of appendicitis in children. J Surg Res 2016;200:452-9.

21. Pogorelić Z, Mihanović J, Ninčević S, Lukšić B, Elezović Baloević S, Polašek O. Validity of appendicitis inflammatory response score in distinguishing perforated from non-perforated appendicitis in children. Children (Basel) 2021;8:309.

22. Stiel C, Elrod J, Klinke M, Herrmann J, Junge CM, Ghabban T, et al. The modified Hedelberg score and the AI appendicitis score are superior than current scores in predicting appendicitis in children: A two -center cohort study. Front Pediatr 2020;8:592892.