

ORIGINAL RESEARCH

Predictive scoring systems to differentiate between simple and complex appendicitis in children

¹Dr. Hraydesh Dixit, ²Dr. Surya Kiran, ³Dr. Aash Sinha, ⁴Dr. Ritik Kathal, ⁵Dr. A K Singh

^{1,2,3,4}Junior Resident, ⁵Professor, Department of General Surgery, Shyam Shah Medical College, Rewa, M.P., India

Corresponding author

Dr. Ritik Kathal

Junior Resident, Department of General Surgery, Shyam Shah Medical College, Rewa, M.P., India

Received: 10 June, 2023

Accepted: 13 July, 2023

ABSTRACT

Background: Several clinical prediction rules have been developed for preoperative differentiation between simple and complex appendicitis in children, as potential treatment strategies differ. This study aimed to externally validate applicable clinical prediction rules that could be used to differentiate between simple and complex appendicitis in children. **Methods:** Potential clinical prediction rules were identified by a scoping review of the literature. Clinical prediction rules applicable in our daily practice were subsequently externally validated in a multicenter historical cohort consisting of 1 tertiary center and 1 large teaching hospital. All children (<18 years old) with histopathologically confirmed acute appendicitis between April 2022 to March 2023 were included. Test results of clinical prediction rules were compared to the gold standard of either simple or complex appendicitis consisting of predefined perioperative and histopathological criteria. Areas under the receiver operating characteristic curves were determined for the selected clinical prediction rules. Areas under the receiver operating characteristic curve >0.7 were considered acceptable and potentially useful. **Results:** In total, 31 clinical prediction rules were identified, of which 12 could be evaluated in our cohort consisting of 550 children. The main reason to exclude clinical prediction rules was the use of variables that were not routinely measured in our cohort. In our cohort, 208/550 (38%) were diagnosed with complex appendicitis according to the gold standard. Clinical prediction rules with areas under the receiver operating characteristic curve >0.7 were: Gorter (0.81), Bogaard (0.79), Bröker (0.79), Graham (0.77), Hansson (0.76), BADCF (0.76), and Eddama (0.75). **Conclusion:** In this study, clinical prediction rules consisting of a combination of clinical and objective variables had the highest discriminative ability. External validation showed that 7 clinical prediction rules were potentially useful. Integration of these clinical prediction rules in daily practice is proposed to guide decision making regarding treatment strategies.

Keywords:

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-Share Alike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

INTRODUCTION

Historically, appendectomy has been standard of care for all children presenting with acute appendicitis. Current international guidelines, however, recommend emergency appendectomy (or drainage procedures in case of appendiceal abscess) for complex appendicitis that should be performed as soon as feasible after diagnosis, versus urgent appendectomy within 24 hours after diagnosis for simple appendicitis.^{1,2,3} Moreover, nonoperative treatment of simple appendicitis has been found to be a safe alternative to appendectomy.^{4,5,6} As potential treatment strategies differ according to appendicitis severity, timely and accurate diagnosis of both simple and complex appendicitis has become increasingly important.

To improve differentiation between both types of appendicitis, several clinical, biochemical, and radiological variables that have been found to be insufficient for differentiation as a standalone modality have been combined into clinical prediction rules (CPRs).^{7,8} Historically, most CPRs have been developed to support the diagnosis of acute appendicitis (without differentiation in type).⁹⁻¹³ Because only recently 2 distinct types of appendicitis are distinguished, few CPRs have been specifically developed to differentiate between simple and complex appendicitis.^{14,15} The studies in which these CPRs were developed showed promising results, but an overview of available CPRs and especially their external validation are lacking.¹⁴⁻¹⁶ It is hypothesized that CPRs are useful to accurately differentiate simple from complex appendicitis and

that, after external validation, CPRs with an acceptable diagnostic accuracy would be identified.

Therefore, the aim of this study was to perform a scoping review of the literature to identify CPRs that differentiate between simple and complex appendicitis and to select CPRs that are applicable in routine daily practice in The Netherlands. Second, we aimed to externally validate the selected CPRs in a multicenter historical cohort of pediatric patients with acute appendicitis.

METHODS

Identification of the CPRs: Scoping review of the literature

A comprehensive search was performed in the PubMed and Embase databases in collaboration with our experienced medical librarian (R.V.) to identify available CPRs. Terms used as index terms or free text words were: “appendicitis,” “predict*,” “score*,” and “decision”. Studies were screened for title and abstract and subsequently assessed for full text by 2 independent reviewers (P.A. and S.T.). Disagreements were resolved by consensus. All CPRs that were tested in previous studies for their potential in differentiating between simple and complex appendicitis were eligible for inclusion. Both CPRs developed in a pediatric population and an adult population were included. CPRs containing variables that are not routinely measured in the participating hospitals (eg, rectal tenderness and biomarkers such as matrix metalloproteinase) and CPRs that used computed tomography (CT) scan results were excluded.

Multicenter historical cohort

In this multicenter historical cohort study, patient records of all children (<18 years old) with histopathologically proven acute appendicitis (simple and complex) treated in Shyam Shah Medical College Rewa MP in department of surgery in between April 2022 to March 2023. Eligible patients were identified using International Classification of Diseases (ICD) codes for acute appendicitis and acute abdomen. Patients who were transferred to either of the participating hospitals after appendectomy elsewhere and those with a finding of a noninflamed appendix during surgery or at histopathological examination were excluded. Furthermore, patients who were initially treated nonoperatively and those who underwent appendectomy for an indication other than acute appendicitis were excluded.

The study protocol was reviewed by both the Medical Ethics Review Committee of the Amsterdam UMC, location AMC, and the Medical Ethics Review Committee of the Northwest hospital. It was confirmed that the Dutch Medical Research Involving Human Subjects Act does not apply and therefore the need for complete ethical review was waived. Patients who objected against the use of their data were

excluded. The results of the study were reported according to the STARD 2015 guidelines.¹⁷

Data extraction

Patient files were reviewed, and data were extracted and stored in an online database (Castor EDC). Two authors independently extracted data (P.A., S.T.), and conflicts were resolved by consensus. Variables were collected to calculate scores of all identified CPRs for each included patient. These collected variables were baseline characteristics (age at presentation, sex, comorbidities), clinical variables (duration of abdominal pain [days], location of pain in right iliac fossa, migration of pain, anorexia, vomiting, intensity of pain [numeric rating scale], progression of pain, constant pain, hopping/coughing/percussion tenderness, tenderness inside/outside right iliac fossa, rebound tenderness, guarding, rigidity, temperature [degrees Celsius], heart rate, micturition difficulties), biochemical variables (leukocytes [$\times 10^9/L$], C-reactive protein [CRP], leukocytes differential count), radiological variables (ultrasonography and magnetic resonance imaging results), perioperative variables (including perioperative diagnosis [simple or complex appendicitis]), and histopathological variables (variables that differentiate between simple and complex appendicitis).

CPR scores were calculated for all patients and compared to our gold standard consisting of a combination of perioperative and histopathological criteria as proposed by Bhangu et al.¹⁸ Simple appendicitis was therefore defined as the macroscopic appearance of an increased diameter of the appendix, without signs of necrosis or perforation, and without the presence of an abscess or mass. Complex appendicitis was defined as appendicitis with signs of necrosis (gangrenous appendicitis) or perforation, or appendicitis with an abscess or mass. Perforation was defined as a (macroscopic or microscopic) visible hole in the appendix or the presence of a free fecalith in the abdominal cavity.

Data analysis

Descriptive analysis of the data was performed. Continuous variables were presented as mean with standard deviation or median with interquartile ranges (IQR) according to their distribution. Scores of all CPRs were calculated for each patient and compared to our gold standard, consisting of a combination of perioperative and histopathological criteria.¹⁸ Complex appendicitis was considered a positive test result in the index test (CPRs) and the gold standard, whereas simple appendicitis was regarded as a negative test result. Subsequently, receiver operating characteristic (ROC) curves were drawn and area under the curve (AUC) values, the primary outcome of this study, were calculated and displayed as proportions with 95% confidence interval (CI). AUC values ≥ 0.7 were considered acceptable and potentially useful. If a CPR could be calculated

for only <50% of our cohort, the CPR was excluded from analysis. If provided, cut-off values as presented in the original manuscript were used to calculate the sensitivity, specificity, predictive values (positive and negative), and likelihood ratios (positive and negative) for these CPRs. These secondary outcomes were displayed as percentages with 95% CI or proportions with 95% CI.

The total cohort was subsequently divided into patients who presented at the Amsterdam University Medical Centers (tertiary referral hospital) and those who presented at the Northwest Hospital Alkmaar (large teaching hospital). Subgroup analyses were performed, and the outcomes as described for the primary analysis were presented. All statistics were performed using IBM SPSS Statistics version 26 (IBM SPSS 26.0, Armonk, NY).

RESULTS

Literature review: Identification of the CPRs

The literature search yielded 6,218 articles, of which 2,477 were found in PubMed and 3,741 in Embase. After removal of duplicates, 4,105 articles were screened for title and abstract. Subsequently, 97 articles were assessed for full text, of which 59 were excluded due to various reasons (Figure 1). The 38 included studies reported on 31 different CPRs that were tested for their ability to differentiate between simple and complex appendicitis. Of these 31 identified CPRs, 14 were excluded. Reasons for exclusion of CPRs were the use of variables that are not routinely measured in our emergency department ($n = 8$) and use of CT scan results ($n = 6$). Variables not routinely measured in our emergency department were rectal tenderness, classification of rebound tenderness into light, medium, and strong, and laboratory markers such as matrix metalloproteinase and erythrocyte sedimentation rate.

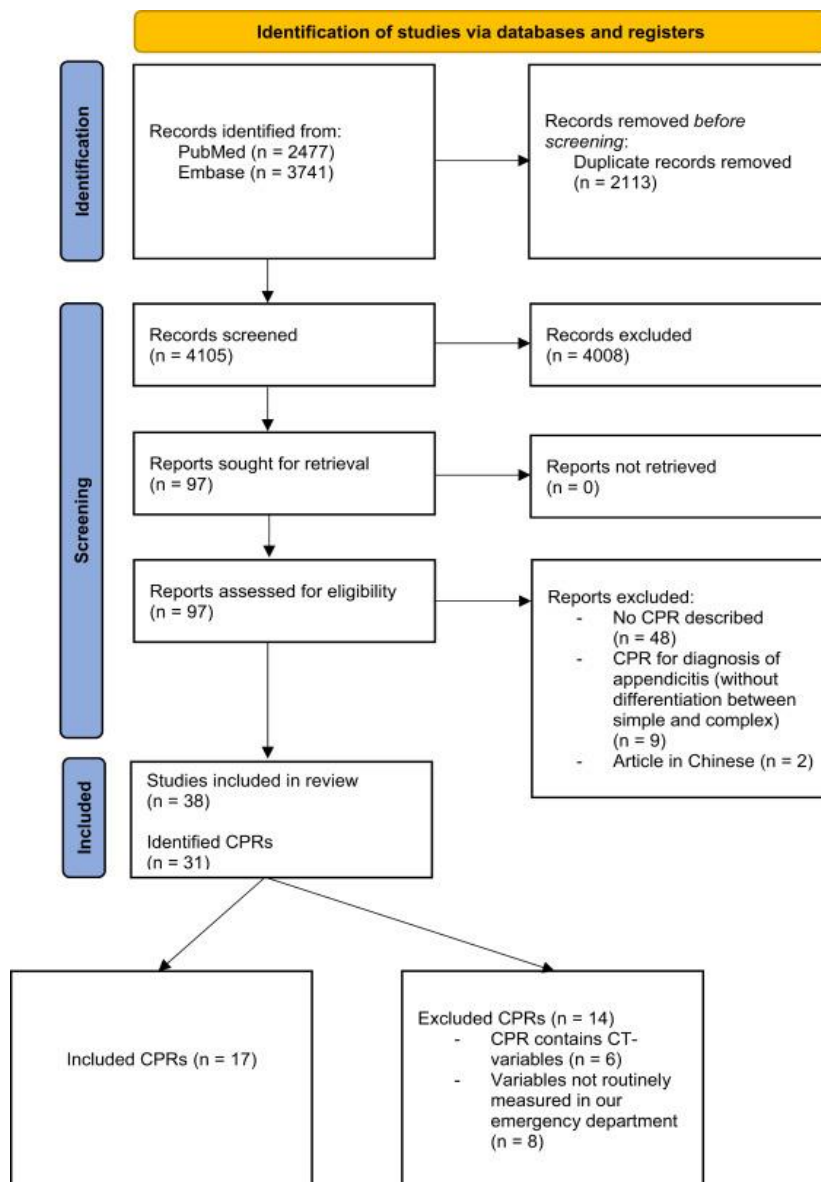


Figure 1. Prisma flowchart literature review. CPR, clinical prediction rule; CT, computed tomography.

The 17 CPRs that were included in this study were: the Alvarado Score, BADCF score, Heidelberg Appendicitis score, Ohmann score, Pediatric Appendicitis Score (PAS), Tzanakis score, and the CPRs developed by Atema, Bogaard, Bonadio, Bröker, Eddama, Feng, Gorter, Graham, Hansson, Kim, and Obinwa. 14, 15, 16-19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32 Characteristics of the CPRs are displayed in Table I and in Supplementary Appendix S3.

Table I. Characteristics of included CPRs

Variables	BAD CF	Boga ard	Brö ker	Edda ma	Gor ter	Grah am	Hans son	H AS	Ki m	Obin wa	Ohm ann	Tzan akis
General												
Age				X			X				X	
Sex				X								
Clinical												
Duration of abdominal pain			X			X						
Constant/increasing pain						X		X			X	
Migration of pain											X	
Location of pain						X						
Anorexia	X									X		
Diarrhea	X											
Micturition difficulties											X	
Heart rate						X						
Temperature	X	X			X	X				X		
Tenderness in RLQ								X			X	X
Abdominal guarding/rebound tenderness						X		X		X	X	X
Rigidity					X						X	
Biochemical												
Leukocytes				X		X	X			X	X	X
CRP	X	X	X	X	X		X		X			
Imaging												
Appendix visualized		X						X				X
Appendix diameter		X							X			
Fat infiltration									X			
Free fluid	X	X							X			
Signs of complex appendicitis					X							

CPR, clinical prediction rule; CRP, C-reactive protein; HAS, Heidelberg Appendicitis Score; RLQ, right lower quadrant.

Multicenter retrospective cohort study

Population and baseline characteristics

During the study period, 630 children were identified using ICD codes for acute appendicitis and acute abdomen. Of these, 80 were excluded for the following reasons: patients were treated nonoperatively ($n = 39$), a noninflamed appendix was found during surgery or at histopathological

examination ($n = 22$), patients were transferred for treatment of complications after appendectomy elsewhere ($n = 14$), patients underwent appendectomy for another indication than acute appendicitis ($n = 3$), and patients objected against the use of their data ($n = 2$). In total, 550 patients with a median age of 11 years (IQR 9–14) were included. Of these patients, 342 (62%) were diagnosed with simple and 208 (38%)

with complex appendicitis according to our gold standard. General characteristics of the study cohort are shown in Table II.

Table II. Baseline characteristics

Characteristics	Missing*	Total (N = 550)
Male (N, %)	0	315 (57)
Age [†]	0	11 (9–14)
Duration of abdominal pain (d) [†]	6	1 (1–2)
Body temperature (°C) [‡]	16	37.5 ± 0.8
Leukocyte count [‡]	8	15.6 ± 5.0
CRP [†]	7	37 (11–91)
Type of appendicitis	0	
Simple (N, %)		342 (62.2)
Complex (N, %)		208 (37.8)
Preoperative imaging procedures	0	
No imaging (N, %)		5 (0.9)
US (N, %)		458 (83.3)
US + MRI (N, %)		81 (14.7)
US + CT (N, %)		2 (0.4)
MRI (N, %)		2 (0.4)
CT (N, %)		2 (0.4)
Type of hospital	0	
Academic (N, %)		107 (19.5)
Large teaching hospital (N, %)		443 (80.5)

Data are displayed as n (% of total).

CRP, C-reactive protein; CT, computed tomography; IQR, interquartile range; MRI, magnetic resonance imaging; US, ultrasound.

CPRs: ROC curves and area under the curves

Table III shows the number of patients for which CPR scores could be calculated and ROC curves were drawn. The CPRs of Atema, Bonadio, Feng, the Alvarado Score, and PAS could not be calculated for at least 50% of patients in our cohort, mainly due to missing data, and were therefore excluded from analysis (Table III). The presence of a fecalith on

imaging and absolute neutrophil count were the most frequently missing variables for the scores of Atema and Bonadio, respectively, and left shift was most frequently missing for both the Alvarado Score and PAS. The CPR of Feng was designed for children up to 5 years old and could therefore not be calculated for at least 50% of patients in this cohort.

Table III. Number of patients for which CPR scores could be calculated

CPR	Total	Amsterdam UMC	Northwest Hospital
Alvarado Score	91/550 (16.5%)	32/107 (29.9%)	59/443 (13.3%)
Atema	235/550 (42.7%)	48/107 (44.9%)	187/443 (42.2%)
BADCF	459/550 (83.5%)	74/107 (69.2%)	385/443 (86.9%)
Bogaard	469/550 (85.3%)	68/107 (63.6%)	401/443 (90.5%)
Bonadio	109/550 (19.8%)	45/107 (42.1%)	64/443 (14.4%)
Bröker	538/550 (97.8%)	98/107 (91.6%)	440/443 (99.3%)
Eddama	539/550 (98.0%)	99/107 (92.5%)	440/443 (99.3%)
Feng*	31/550 (5.6%)	21/107 (19.6%)	10/443 (2.3%)
Gorter	503/550 (91.5%)	81/107 (75.7%)	422/443 (95.3%)
Graham	371/550 (67.5%)	56/107 (52.3%)	315/443 (71.1%)
Hansson	524/550 (95.3%)	96/107 (89.7%)	428/443 (96.6%)
HAS	477/550 (86.7%)	66/107 (61.7%)	411/443 (92.8%)
Kim	448/550 (81.5%)	66/107 (61.7%)	382/443 (86.2%)
Obinwa	485/550 (88.2%)	78/107 (72.9%)	407/443 (91.9%)
Ohmann Score	427/550 (77.6%)	72/107 (67.3%)	355/443 (80.1%)
PAS	85/550 (15.5%)	32/107 (29.9%)	53/443 (12.0%)
Tzanakis	511/550 (92.9%)	86/107 (80.4%)	433/443 (97.7%)

Data are displayed as n/N (% of total).

CPR, clinical prediction rule, PAS, Pediatric Appendicitis Score, UMC, University Medical Centre.

CPR developed for only patients younger than 5 years old.

The ROC curves of the CPRs are shown in the Supplementary Appendix S2, and the corresponding AUC values are presented in Table IV. In our cohort, CPRs with an AUC value of >0.7 were the CPRs of Gorter (0.81), Bogaard (0.79), Bröker (0.79), Graham (0.77), Eddama (0.76), BADCF score (0.76), and Hansson (0.75). Subgroup analysis of the performance of CPRs for each hospital individually showed that the CPRs of Bogaard (AUC value: 0.90), Bröker (AUC value: 0.86), Eddama (AUC value:

0.81), Gorter (AUC value: 0.88), BADCF score (AUC value: 0.77), Hansson (AUC value: 0.76), and Kim (AUC value: 0.76) had an acceptable discriminative power for patients who presented at the Amsterdam UMC (tertiary referral center) (Supplementary Appendix S4). In the cohort of patients who presented at the Northwest Hospital (large peripheral teaching hospital), the CPRs of Bogaard (0.76), Bröker (0.76), Eddama (0.76), Gorter (0.78), Graham (0.78), Hansson (0.78), and BADCF score (0.75) had an AUC value >0.7 .

Table IV. Diagnostic accuracy of CPRs (total cohort)

CPR	Cut-off value	AUC	Sensitivity	Specificity	PPV	NPV	PLR*	NLR*
BADCF	≥ 5	0.76 (0.72–0.81)	90.2% (84.5–94.0%)	40.9% (35.2–46.9%)	48.0% (42.5–53.6%)	87.3% (80.2–92.2%)	1.5 (1.4–1.7)	0.2 (0.2–0.4)
Bogaard	≥ 0.15	0.79 (0.74–0.83)	70.5% (62.8–77.2%)	72.3% (66.8–77.2%)	58.2% (51.0–65.0%)	81.7% (76.4–86.1%)	2.5 (2.1–3.1)	0.4 (0.3–0.5)
Bröker [†]	-	0.79 (0.75–0.83)	-	-	-	-	-	-
Eddama [†]	-	0.76 (0.72–0.80)	-	-	-	-	-	-
Gorter	≥ 4	0.81 (0.77–0.85)	67.0% (59.7–73.6%)	80.3% (75.4–84.5%)	67.0% (59.7–73.6%)	80.3% (75.4–84.5%)	3.4 (2.7–4.3)	0.4 (0.3–0.5)
Graham	≥ 3	0.77 (0.72–0.82)	92.1% (85.9–95.8%)	34.5% (28.5–41.0%)	45.7% (39.8–51.7%)	87.9% (79.0–93.5%)	1.4 (1.3–1.6)	0.2 (0.1–0.4)
Hansson [†]	-	0.75 (0.71–0.80)	-	-	-	-	-	-
HAS	≥ 3	0.48 (0.43–0.54)	84.9% (78.4–89.8%)	12.2% (8.9–16.5%)	34.1% (29.5–38.9%)	60.3% (47.2–72.2%)	1.0 (0.9–1.0)	1.2 (0.8–1.9)
Kim	≥ 2	0.69 (0.64–0.74)	98.1% (94.2–99.5%)	12.2% (8.8–16.7%)	38.5% (33.8–43.5%)	92.1% (77.5–97.9%)	1.1 (1.1–1.2)	0.2 (0–0.5)
Obinwa [†]	-	0.66 (0.61–0.71)	-	-	-	-	-	-
Ohmann [†]	-	0.53 (0.48–0.59)	-	-	-	-	-	-
Tzanakis	≥ 12.5	0.54 (0.49–0.59)	43.7% (36.7–50.9%)	58.7% (53.1–64.1%)	39.3% (32.8–46.1%)	63.0% (57.2–68.4%)	1.1 (0.9–1.3)	1.0 (0.8–1.1)

Data are displayed as percentage (95% confidence interval).

AUC, area under the curve; CPR, clinical prediction rule; NLR, negative likelihood ratio; NPV, negative predictive value; PLR, positive likelihood ratio; PPV, positive predictive value.

CPRs: Secondary outcomes

The diagnostic accuracy of the CPRs, according to the cut-off values as provided by the original articles, is displayed in Table IV, and the STARD flow diagrams can be found in the Supplementary Appendix S5. For 7 out of the 12 identified CPRs, a cut-off value was

provided by the original article. The negative predictive values of the BADCF score, the CPRs of Bogaard, Gorter, Graham, and Kim with their original cut-off values are $\geq 80\%$, implicating that $\geq 80\%$ of patients with a test result of simple appendicitis were also diagnosed with simple appendicitis according to

the gold standard. Focusing on ruling out complex appendicitis, in our cohort the BADCF score and the CPRs of Graham and Kim had the best negative likelihood ratios (0.2, 95% CI 0.2–0.4, 0.1–0.4 and 0–0.5, respectively).

The CPR of Gorter has the highest positive likelihood ratio (3.4, 95% CI 2.7–4.3), although still approximately one-third of patients with complex appendicitis according to the CPR were diagnosed with simple appendicitis according to the gold standard.

DISCUSSION

This study identified 31 CPRs through a scoping review of the literature, of which 17 were included for external validation. Owing to missing data, only 12 of these CPRs were included in our final analysis. External validation of these CPRs in our multicenter historical cohort showed that 7 (Gorter, Bogaard, Bröker, Graham, Eddama, BADCF, and Hansson) had an AUC value ≥ 0.7 and thus were considered as potentially useful in our daily practice.^{15,16,22,24,25,31,33}

Accurate preoperative differentiation between simple and complex appendicitis has become increasingly important, as potential treatment strategies for both types of appendicitis differ. International guidelines on the management of acute appendicitis recommend emergency appendectomy (as soon as feasible after diagnosis) for complex appendicitis, because this strategy is associated with better outcomes.^{1, 2, 3} Initial nonoperative treatment of simple appendicitis is found to be safe and effective and therefore has become a potential alternative for appendectomy. When appendectomy is preferred for simple appendicitis, it can safely be delayed up to 24 hours after diagnosis, since several studies have shown that a short time delay from emergency department to surgery did not lead to increased appendiceal perforation rates or worse outcomes.^{3,6,34,35} These guidelines were formed by international societies and are followed in many Western countries.^{1,2}

Thus far, the ability of several variables to differentiate between simple and complex appendicitis has been tested and some clinical variables, such as duration of abdominal pain, have been found to be predictive for complex appendicitis.^{36, 37, 38} However, most clinical variables derived from history and physical examination are subject to interpretation bias, which reduces their reproducibility.³⁹ Therefore, the potential discriminative ability of (objective) biomarkers has also been investigated, demonstrating high diagnostic accuracy of white blood cell count, granulocyte count, and especially CRP for prediction of complex appendicitis.^{7,36,40, 41, 42} It is hypothesized that the discriminative ability of biomarkers could be improved in combination with other objective predictors, such as imaging modalities. Radiological variables highly specific for appendiceal

perforation are signs of an abscess, intraluminal fecalith, and especially loss of the submucosal layer of the appendix.^{43, 44, 45} Although these variables are potentially useful, the overall diagnostic accuracy of ultrasound as a standalone modality is insufficient.⁸ Therefore, it has been suggested that combining these objective variables into CPRs would optimize differentiation between simple and complex appendicitis. This was confirmed in the present study, as the CPRs that combine objective variables such as vital parameters and laboratory results with radiological variables (Gorter, Bogaard, BADCF) had the highest AUC values. Especially objective variables such as body temperature, CRP, and free fluid on imaging are included in the majority of these CPRs and are therefore found to be highly discriminative. In contrast, CPRs mainly consisting of physical examination variables (Heidelberg Appendicitis Score and Ohmann) were the least useful for differentiation between simple and complex appendicitis.

This study identified 7 CPRs, with a wide variety of clinical, biochemical, and radiological variables. These CPRs could be used for clinical decision-making about timing of appendectomy (emergency appendectomy versus urgent appendectomy within 24 hours after diagnosis) or for selecting patients who are eligible for nonoperative treatment. Selecting a CPR for use in daily clinical practice depends on 2 main factors. First, it depends on whether the integrated variables in the CPRs are already routinely assessed in daily practice and, second, on the purpose of integrating CPRs in clinical practice. Of note, some CPRs are more useful for selecting patients for nonoperative treatment, whereas others are more accurate in identifying patients with complex appendicitis in need of emergency treatment. For the former purpose, a CPR with a high sensitivity should be chosen to ensure that no patients with complex appendicitis are selected for nonoperative treatment. For the latter purpose, however, CPR with a high specificity should be selected to ensure prompt treatment for those patients with complex appendicitis. In our daily practice, the CPR of Gorter is used in a randomized controlled trial for selection of patients who are eligible for nonoperative treatment.⁴⁶ This CPR was chosen because of the high sensitivity and negative predictive value that were found in a previous validation study, and because it consists of 5 variables that are routinely assessed in our daily clinical practice.¹⁵

Difference in performance of the CPRs in the original studies compared to the present study can first be explained by the fact that most of the CPRs tested in our study were developed and validated in the same cohort of patients and therefore lack external validation. The performance of a CPR is generally lower in cohorts of patients other than the one in which it was developed. This could be one of the main reasons that the AUC values of the selected CPRs

were lower in the present study compared to the original studies. Only the CPR of Gorter was externally validated after development, yielding a sensitivity of 86%, specificity of 91%, positive predictive value of 67%, negative predictive value of 97%, positive likelihood ratio of 10, and negative likelihood ratio of 0.15.¹⁵ The present study found a slightly lower diagnostic accuracy, which could be explained by differences in population. The original CPR was validated in a peripheral teaching hospital, whereas our cohort consisted of patients from both a tertiary referral center and a peripheral teaching hospital. For the development of future CPRs, external validation is recommended. Second, differences could be explained by the fact that the goals for which the CPRs were developed vary. Some CPRs were specifically developed for the differentiation between simple and complex appendicitis,^{15,22,24,31} whereas others were originally developed for the diagnosis of acute appendicitis (without differentiation of type) and were tested for their ability of differentiating between simple and complex appendicitis in previous studies.^{27,29,47} Third, the present study found a higher discriminative ability of CPRs in the tertiary center subgroup compared to the peripheral teaching hospital, which could be explained by different characteristics of the subgroups. In general, children treated at the tertiary center were younger, and a higher proportion of patients were diagnosed with complex appendicitis compared to the patients treated at the peripheral teaching hospital.

The results of our study should be interpreted in the light of some limitations. Data regarding the variables of the CPRs and outcomes were retrospectively collected, possibly leading to information bias. Due to missing data, 5 of 17 identified CPRs could not be calculated for at least 50% of patients in our cohort and were therefore excluded from analysis. Moreover, 14 out of 31 CPRs were excluded because they contained variables that were not routinely collected in our daily practice. Therefore, these CPRs could not be tested in our cohort of patients, although they could potentially be useful in another population. Furthermore, for the CPRs that used ultrasound variables, ultrasound reports were not reassessed by a radiologist. Moreover, ultrasound performance and conclusions are prone to interobserver variability, which is also dependent on the skills of the radiologist. Additionally, the patients included in our study were identified using ICD codes for acute appendicitis and acute abdomen, which might have led to selection bias in cases of wrong ICD code classification. Moreover, as acute appendicitis is a disease with different behaviors in different pediatric age groups (eg, children younger than 5 years old versus teenagers), it could be interesting to investigate the applicability of the CPRs in different age groups. Lastly, generalizability of the results of this study could be questioned, as those CPRs that are applicable

in our daily practice were selected. However, nowadays, global guidelines on the diagnostic work-up and treatment of acute appendicitis are followed by many countries; therefore, management of patients is becoming increasingly comparable.^{1,2} Furthermore, in this study, 7 CPRs consisting of different clinical, biochemical, and radiological variables have been identified as potentially useful. Therefore, surgeons from different countries can select the CPR that consists of variables that are routinely assessed in their daily practice, thereby keeping in mind the purpose for which they are planning to use the CPR. In the present study, the CPRs with the best discriminative ability were those developed in a population consisting of patients treated in The Netherlands.^{15,31,33} These CPRs were therefore tested in a patient cohort that was relatively similar to the population in which the CPRs were developed, which could have attributed to the high diagnostic accuracy. Thus, apart from selecting one of the CPRs identified in the present study, it might also be beneficial for clinicians to develop a CPR in their own population. Strengths of this study include the predefined definitions for simple and complex appendicitis based on the perioperative and histopathological classification as proposed by Bhangu et al.¹⁸ This resulted in an objective and reproducible classification of appendicitis severity according to the gold standard. Furthermore, all data were collected by one author, and, subsequently, the entire database was checked by another author. Classification of simple and complex appendicitis according to the gold standard was also independently performed by 2 authors. Additionally, our cohort consisted of patients treated in a tertiary referral center and a large peripheral teaching hospital, which improved the generalizability of our results. This is also reflected by the distribution of simple versus complex appendicitis in this study, which was comparable to the distribution that was found in a recent nationwide audit in The Netherlands.

CONCLUSION

External validation of 12 CPRs in a multicenter retrospective cohort showed that 7 CPRs had an AUC value >0.7 and were therefore potentially useful for the differentiation between simple and complex appendicitis in our population. CPRs consisting of a combination of clinical, biochemical, and radiological variables were found to have the highest discriminative ability. We propose the integration of these 7 CPRs in daily clinical practice to guide decision making regarding treatment strategies for children with acute appendicitis.

REFERENCES

1. S. Di Saverio, M. Podda, B. De Simone, et al. Diagnosis and treatment of acute appendicitis: 2020 update of the WSES Jerusalem guidelines World J Emerg Surg, 15 (2020), p. 27

2. R.R. Gorter, H.H. Eker, M.A. Gorter-Stam, et al. Diagnosis and management of acute appendicitis. EAES consensus development conference 2015 *Surg Endosc*, 30 (2016), pp. 4668-4690
3. Dutch Association for Surgery Dutch national guideline: diagnosis and treatment of acute appendicitis *Federatie Medisch Specialisten* (2019)
4. M. Knaapen, J.H. van der Lee, H.A. Heij, E.L.W. van Heurn, R. Bakx, R.R. Gorter. Clinical recovery in children with uncomplicated appendicitis undergoing non-operative treatment: secondary analysis of a prospective cohort study *Eur J Pediatr*, 178 (2019), pp. 235-242
5. E.H. Livingston, T.B. Fomby, W.A. Woodward, R.W. Haley. Epidemiological similarities between appendicitis and diverticulitis suggesting a common underlying pathogenesis *Arch Surg*, 146 (2011), pp. 308-314.
6. P.C. Minneci, E.M. Hade, A.E. Lawrence, et al. Association of nonoperative management using antibiotic therapy vs laparoscopic appendectomy with treatment success and disability days in children with uncomplicated appendicitis *JAMA*, 324 (2020), pp. 581-593
7. R.E. Andersson. Meta-analysis of the clinical and laboratory diagnosis of appendicitis *Br J Surg*, 91 (2004), pp. 28-37
8. D.J. Nijssen, P. van Amstel, J. van Schuppen, L.D. Eeftinck Schattenkerk, R.R. Gorter, R. Bakx. Accuracy of ultrasonography for differentiating between simple and complex appendicitis in children. *Pediatr Surg Int*, 37 (2021), pp. 843-849
9. Alvarado. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med*, 15 (1986), pp. 557-564
10. M. Andersson, R.E. Andersson. The appendicitis inflammatory response score: a tool for the diagnosis of acute appendicitis that outperforms the Alvarado score *World J Surg*, 32 (2008), pp. 1843-1849
11. M.H. Ebell, J. Shinholser. What are the most clinically useful cutoffs for the Alvarado and Pediatric Appendicitis Scores? A systematic review. *Ann Emerg Med*, 64 (2014), pp. 365-372.e2
12. D.M. Kulik, E.M. Uleryk, J.L. Maguire. Does this child have appendicitis? A systematic review of clinical prediction rules for children with acute abdominal pain. *J Clin Epidemiol*, 66 (2013), pp. 95-104
13. P. van Amstel, R.R. Gorter, J.H. van der Lee, H.A. Cense, R. Bakx, H.A. Heij. Ruling out appendicitis in children: can we use clinical prediction rules? *J Gastrointest Surg*, 23 (2019), pp. 2027-2048
14. J.J. Atema, C.C. van Rossem, M.M. Leeuwenburgh, J. Stoker, M.A. Boermeester. Scoring system to distinguish uncomplicated from complicated acute appendicitis. *Br J Surg*, 102 (2015), pp. 979-990
15. R.R. Gorter, A.L. van den Boom, H.A. Heij, et al. A scoring system to predict the severity of appendicitis in children. *J Surg Res*, 200 (2016), pp. 452-459
16. Y. Lee, H. Cho, G. Gwak, B. Bae, K. Yang. Scoring system for differentiation of complicated appendicitis in pediatric patients: appendicitis scoring system in children. *Glob Pediatr Health*, 8 (2021) 2333794X2111022268
17. P.M. Bossuyt, J.B. Reitsma, D.E. Bruns, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies *BMJ*, 351 (2015), p. h5527
18. Bhangu, K. Søreide, S. Di Saverio, J.H. Assarsson, F.T. Drake. Acute appendicitis: modern understanding of pathogenesis, diagnosis, and management *Lancet*, 386 (2015), pp. 1278-1287
19. O.O. Adibe, O.J. Muensterer, K.E. Georgeson, C.M. Harmon. Severity of appendicitis correlates with the pediatric appendicitis score *Pediatr Surg Int*, 27 (2011), pp. 655-658
20. M. Boettcher, P. Günther, T. Breil. The Heidelberg Appendicitis Score predicts perforated appendicitis in children. *Clin Pediatr (Phila)*, 56 (2017), pp. 1115-1119
21. W. Bonadio, S. Shahid, L. Vardi, et al. A pre-operative clinical scoring system to distinguish perforation risk with pediatric appendicitis *J Pediatr Surg*, 53 (2018), pp. 441-445
22. M. Eddama, K.C. Fragkos, S. Renshaw, et al. Logistic regression model to predict acute uncomplicated and complicated appendicitis *Ann R Coll Surg Engl*, 101 (2019), pp. 107-118
23. W. Feng, X.F. Zhao, M.M. Li, H.L. Cui. A clinical prediction model for complicated appendicitis in children younger than five years of age. *BMC Pediatr*, 20 (2020), p. 401
24. D.F. Graham. Computer-aided prediction of gangrenous and perforating appendicitis. *BMJ* (1977), pp. 1375-1377
25. J. Hansson, A. Khorram-Manesh, A. Alwindawe, K. Lundholm. A model to select patients who may benefit from antibiotic therapy as the first line treatment of acute appendicitis at high probability. *J Gastrointest Surg*, 18 (2014), pp. 961-967
26. T.H. Kim, B.S. Cho, J.H. Jung, M.S. Lee, J.H. Jang, C.N. Kim. Predictive factors to distinguish between patients with noncomplicated appendicitis and those with complicated appendicitis. *Ann Coloproctol*, 31 (2015), pp. 192-197
27. M. Nissen, R.B. Trobs. Levels of C-reactive protein and sodium may differentiate a perforated appendix from a nonperforated appendix in children. *Biomed Res Int*, 2021 (2021), p. 9957829
28. O. Obinwa, C. Peirce, M. Cassidy, T. Fahey, J. Flynn. A model predicting perforation and complications in paediatric appendicectomy. *Int J Colorectal Dis*, 30 (2015), pp. 559-565
29. Ohmann, C. Franke, Q. Yang. Clinical benefit of a diagnostic score for appendicitis: results of a prospective interventional study. *German Study Group of Acute Abdominal Pain. Arch Surg*, 134 (1999), pp. 993-996
30. CFd Sousa-Rodrigues, ACd Rocha, A.K.B. Rodrigues, F.T. Barbosa, FWdS. Ramos, S.H.C. Valões. Correlation between the Alvarado Scale and the macroscopic aspect of the appendix in patients with appendicitis. *Revista do Colégio Brasileiro de Cirurgiões*, 41 (2014), pp. 336-339
31. V.A. van den Bogaard, S.M. Euser, T. van der Ploeg, et al. Diagnosing perforated appendicitis in pediatric patients: a new model. *J Pediatr Surg*, 51 (2016), pp. 444-448
32. E.M. Yilmaz, M. Kapci, S. Celik, B. Manoglu, M. Avcil, E. Karacan. Should Alvarado and Ohmann scores be real indicators for diagnosis of appendicitis and severity of inflammation? *Ulus Travma Acil Cerrahi Derg*, 23 (2017), pp. 29-33
33. M.E. Bröker, E.M. van Lieshout, M. van der Elst, L.P. Stassen, T. Schepers. Discriminating between simple

- and perforated appendicitis. *J Surg Res*, 176 (2012), pp. 79-83
34. L. Huang, Y. Yin, L. Yang, C. Wang, Y. Li, Z. Zhou. Comparison of antibiotic therapy and appendectomy for acute uncomplicated appendicitis in children: a meta-analysis. *JAMA Pediatr*, 171 (2017), pp. 426-434
 35. M.D. Stevenson, P.S. Dayan, N.C. Dudley, et al. Time from emergency department evaluation to operation and appendiceal perforation *Pediatrics*, 139 (2017)
 36. H. Miyauchi, Y. Okata, T. Hatakeyama, T. Nakatani, Y. Nakai, Y. Bitoh. Analysis of predictive factors for perforated appendicitis in children. *Pediatr Int*, 62 (2020), pp. 711-715