Prognostic capabilities and agreement of three different scores in diagnosing appendicitis in children from a developing setting

Hernando Pinzón-Redondo[a,b,c], Andrea Zarate-Vergara[a], Katherine Barrios-Redondo[a], Cesar Muñoz[a], Ángel Guzmán[a], Dorys Morales-Payares[a], Nelson Alvis-Guzmán[a,b] and Ángel Paternina-Caicedo[a,b]

Objective **The aim of this study was to assess the prognostic capabilities of the Pediatric Appendicitis Score (PAS), the Alvarado score, and the modified Alvarado score in diagnosing appendicitis in children in a developing country setting (Cartagena, Colombia).**

Methods **A retrospective cohort study was designed and carried out in Hospital Infantil Napoleón Franco Pareja (HINFP), a reference pediatric center in Cartagena, Colombia. Children under 18 years of age, of any sex, admitted during 2013 were included in the present study. The sensitivity and specificity of each score value in diagnosing appendicitis were estimated through the area under the receiver operative characteristics curve (AUC).**

Results **A total 187 patients were admitted in HINFP with abdominal pain presumptive of appendicitis. The median age of the cohort sample was 11.58 years (interquartile range, 8.33–13.61); AUC was not statistically different in the three scores assessed ($P=0.549$). PAS had an AUC of 0.628 [95% confidence interval (CI), 0.495–0.763]; Alvarado score had an AUC of 0.642 (95% CI, 0.514–0.770); and modified Alvarado score had an AUC of 0.611 (95% CI, 0.471–0.751).**

Conclusion **Our study shows a relatively poor overall performance of the scores in diagnosing appendicitis in a developing income setting. The applicability and prognostic usefulness of the PAS, the Alvarado score, and the modified Alvarado score in developing countries should be based on the stratification of appendicitis risk in the pediatric population.** ***Ann Pediatr Surg*** **12:5–9 © 2016 Annals of Pediatric Surgery.**

Annals of Pediatric Surgery 2016, **12**:5–9

Keywords: appendicitis, developing countries, diagnosis, sensitivity, specificity

[a]Grupo de Investigación Infectología Pediátrica, Hospital Infantil Napoleón Franco Pareja, [b]Grupo de Investigación en Economia de la Salud – GIES and [c]Departamento de Pediatria, Facultad de Medicina, Universidad de Cartagena, Cartagena, Colombia

Correspondence to Ángel Paternina-Caicedo, MD, MSc, Centro de Investigación y Docencia – CID, Hospital Infantil Napoleón Franco Pareja 'Casa del Niño', Brr Bruselas, Cartagena, Bolívar 130001, Colombia
Tel: +57 5 6535046; fax: +57 5 6535046;
e-mails: angel.paternina@gmail.com, ajpaternac@unal.edu.co

Received 4 July 2015 **accepted** 7 December 2015

Introduction

Appendicitis is the most common cause of abdominal surgery in children, with an estimated annual incidence rate of 26 appendectomies per 100 000 population [1]. The diagnosis of appendicitis continues to be mostly based on clinical symptomatology and natural history of the disease. Several scores have been designed and validated in children to aid in management decisions for patients with appendicitis. The Pediatric Appendicitis Score (PAS) [2], the Alvarado score, the modified Alvarado score [3], the Kharbanda score [4], the Lintula score [5], and the Van den Broek scores [6] were designed for use in pediatric population, but validation has provided mixed results of their diagnostic utility [7].

The validation of all these scores had been made previously in developed settings, proving useful in some instances for decision making in children with appendicitis. However, in developing settings, where case mix and comorbidities or baseline characteristics can be different, some variables may impact the diagnostic usefulness of these scores. In children from developing countries, for example, stunting and age of presentation may impact symptom presentation, delay attention and treatment, and affect the overall score performance. In Colombia, for example, an upper-middle income country in South America, stunting frequency in children is 15% [8].

The diagnostic utility of PAS, the modified Alvarado score, and the Alvarado score, the most used and researched scores for diagnosing appendicitis, has been tested infrequently in a low-income and middle-income setting population. We therefore carried out a cohort study with the aim of testing the diagnostic capacity of these scores in diagnosing appendicitis in children from a developing country center from Colombia.

Methods

A retrospective cohort study was designed and carried out to test the diagnostic capacity of the PAS and the Alvarado score in diagnosing appendicitis in children from Cartagena (Colombia). The setting of the present assessment is Hospital Infantil Napoleón Franco Pareja (HINFP), a reference pediatric center in a city with approximately one million inhabitants.

Children presenting to the emergency department at HINFP in 2013 (1 January to 31 December), with unspecified abdominal pain suggestive of appendicitis as

1687-4137 © 2016 Annals of Pediatric Surgery DOI: 10.1097/01.XPS.0000476033.40918.0f

the main complaint, under 18 years of age and of any sex, were included in the present study.

Variable definitions

Population characteristics

Variables collected for this study were age, weight, sex, rural residence, key vital signs at admission (pulse, respiratory rate, and temperature), white blood cell count, percentage of neutrophils, hospitalization duration, and ICU stay.

Gold standard

True disease was defined as histopathology from appendectomy suggestive of appendicitis. The gold standard also included ruling out appendicitis at the time of discharge after a complete follow-up.

Pediatric Appendicitis Score

The PAS is an eight-item pediatric score for predicting appendicitis [2]. The full list of symptomatology of this score is listed in Table 1.

Alvarado score

Alvarado [9] designed an eight-item pediatric score for predicting appendicitis. The full list of signs and symptoms is also listed in Table 1.

Modified Alvarado score

This score only differs from the Alvarado score in the absence of neutrophilia as a predictor of appendicitis [3].

Data collection and procedures

Health personnel in the emergency department at our institution follows specific protocols to assess patients with suspected appendicitis. At first contact with physicians, the health personnel retrospectively collected, if present, information on the following: cough, percussion, heel tapping tenderness at interquartile range (IQR)/rebound pain; anorexia; migration of pain to IQR; nausea/vomit; IQR tenderness on light palpation; white blood cell count; neutrophil count; and temperature. All data were retrospectively collected at emergency admission, including the required laboratory data, from electronic health records.

Data analysis

All analysis assumed a *P*-value less than 0.05 as statistically significant, and were carried out using Stata (Stata v.13; StataCorp, College Station, Texas, USA).

Categorical variables were reported as percentages. Continuous variables were reported as mean or median, depending on variable normality. Dispersion measures for continuous variables were SDs, and 25th and 75th percentiles IQR, depending on the normality or non-normality of the variable, respectively. Normality was assessed using the Shapiro–Wilk test for each continuous variable.

Analyses of categorical variables were performed with the χ^2 or Fisher exact test, when appropriate. Analysis of variance or the Kruskal–Wallis rank test were used for analysis of continuous variables, according to the parameter distribution.

The sensitivity and specificity of each score value in diagnosing appendicitis were estimated through the receiver operative characteristics curve results, and 95% confidence intervals (CI) were estimated. Area under the receiver operative characteristic curve (AUC) with 95% CI was also reported. We excluded patients with missing data in any variable of the analysis.

Sensitivity analysis

To account for differences in performance related to age, a sensitivity analysis was performed including only patients older than 5 years of age. AUC with 95% CI and each score value were reported.

Results

Patient characteristics

A total of 236 patients were admitted to HINFP with abdominal pain presumptive of appendicitis. Of them, 49 patients (20.8%) had missing data, and hence were excluded from the study. The median age of the cohort sample was 11.58 years (IQR, 8.33–13.61), and 22 of 187 (11.8%) patients were under 5 years of age. The median hospitalization time was 4 days (IQR, 3–6); six (3.2%) patients were admitted to the ICU during their first hospitalization, for a median length of ICU stay of 6 days (IQR, 4–8). The median stay in the ICU was not different in patients with appendicitis when compared with nonappendicitis patients ($P = 0.164$).

A more comprehensive overview of sample characteristics is shown in Table 2. According to our analysis, none of the vital signs recorded at admission showed differences among appendicitis versus nonappendicitis patients (pulse, respiratory rate, and temperature).

Prognostic capabilities of the scores

With regard to the symptomatology and parameters of the three scores, only neutrophilia was associated with appendicitis in our sample ($P = 0.008$). Below, a detailed overview of the prognostic ability of the scores is given. AUC was not statistically different in the three scores assessed ($P = 0.549$) (Table 3).

Receiver operator curve of the Alvarado score, the modified Alvarado score, and the PAS in our setting is shown in Fig. 1. The percentage of patients per score-value in each score for patients with and without appendicitis in the sample is listed in Fig. 2, and its sensitivity and specificity in Table 4.

Pediatric Appendicitis Score

Twelve (54.6%) children without appendicitis had an Alvarado score less than 6, and 117 (71.3%) appendicitis patients had a score of 6 or greater ($P = 0.109$). The median PAS score was 6 (IQR, 4–7) in patients without appendicitis versus 7 (IQR, 5–7) in patients with appendicitis ($P = 0.050$). AUC was 0.628 (95% CI, 0.495–0.763).

Table 1 Scoring systems for evaluating appendicitis, as reported in the literature

Alvarado score	P	Modified Alvarado score	P	PAS	P
Migration of pain	1	Migration of pain	1	Migration of pain	1
Anorexia/acetone	1	Anorexia/acetone	1	Anorexia	1
Nausea/vomiting	1	Nausea/vomiting	1	Nausea/vomiting	1
IQR tenderness	2	IQR tenderness	2	IQR tenderness	2
Rebound pain	1	Rebound pain	1	Cough/hopping/percussion tenderness in the IQR	2
Elevation in temperature	1	Elevation in temperature	1	Elevation in temperature	1
Leukocytosis	2	Leukocytosis	2	Leukocytosis	1
Differential WBC count with left shift	1	–	–	Differential WBC count with a left shift	1

P, number of points if present; PAS, Pediatric Appendicitis Score; WBC, white blood cell.

Table 2 Sample characteristics of patients with suspected appendicitis at emergency admission in HINFP (Cartagena, Colombia)

Variables	Not appendicitis ($n=22$) (11.8%)	Appendicitis ($n=164$) (88.2%)	Total ($n=186$) (100%)	*P*-value
Characteristics				
Age [median (IQR)] (years)	10.75 (4.51–14.12)	11.63 (8.41–13.59)	11.58 (8.33–13.61)	0.848
Weight [median (IQR)] (kg)	25.5 (13–53)	35 (25–46)	34 (24–47)	0.206
Male sex [*n* (%)]	16 (72.7)	103 (62.8)	119 (64.0)	0.363
Rural precedence [*n* (%)]	7 (33.3)	36 (22.6)	43 (23.9)	0.280
Key vital signs at admission [median (IQR)]				
Heart rate (pulse/min)	96 (87–115)	96 (87.5–110)	96 (87–110)	
Respiratory rate (breaths/min)	23.5 (20–33)	22 (20–28)	22 (20–28)	0.985
Temperature (°C)	37 (37–37.7)	37 (36.95–37.5)	37 (37–37.5)	0.207
White blood cell [median (IQR)]				
White blood cells count (10^3/μl)	13.3 (8.7–17.2)	15.9 (11.9–20.0)	15.5 (11.5–20.0)	0.112
% of neutrophils	71 (61–78)	80.5 (72–86)	79.5 (71–85)	<0.001
Hospitalization time				
Hospitalization days [median (IQR)]	4.5 (2–7)	4 (3–6)	4 (3–6)	0.971
ICU [*n* (%)]	2 (9.1)	4 (2.4)	6 (3.2)	0.149[a]
ICU days [median (IQR)]	8.5 (7–10)	4.5 (3.5–6.5)	6 (4–8)	0.164

ICU, intensive care unit; IQR, interquartile range.
[a]Fisher exact test.

Table 3 Signs and symptoms used in Pediatric Appendicitis Score, Alvarado score, and modified Alvarado score to diagnose appendicitis

Symptomatology	Not appendicitis ($n=22$) (11.8%)	Appendicitis ($n=164$) (88.2%)
Cough/percussion/heel tapping tenderness at IQR/rebound pain	14 (63.6)	128 (78.0)
Anorexia	2 (9.1)	20 (12.2)
Migration of pain to IQR	8 (36.4)	59 (36.0)
Nausea/vomit	17 (77.3)	130 (79.3)
IQR tenderness on light palpation	17 (77.3)	145 (88.4)
Neutrophilia	9 (40.9)	114 (69.5)
Leukocytosis	16 (72.7)	140 (85.4)
Temperature > 38°C (PAS)	3 (13.6)	22 (13.4)
Temperature > 37.3°C (Alvarado and modified Alvarado)	7 (31.8)	48 (29.3)

Definition of leukocytosis was a blood cell count greater than 10 000/ml, and neutrophilia was a left shift > 75%.
PAS, Pediatric Appendicitis Score.

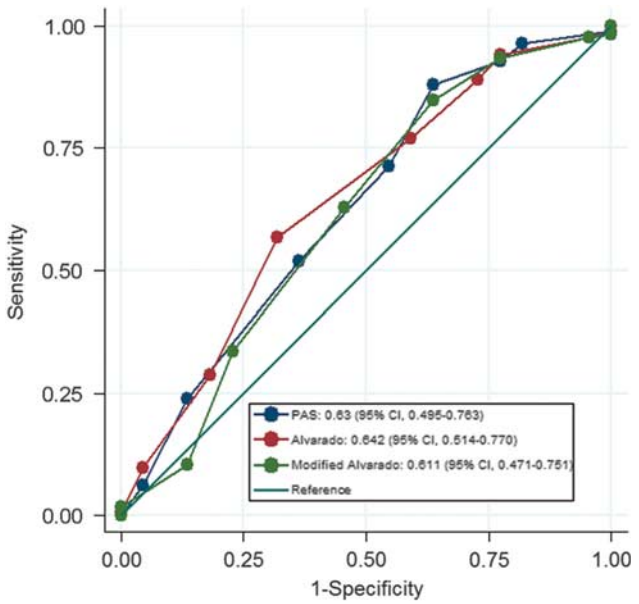
Alvarado score

Seven (31.8%) children without appendicitis had an Alvarado score less than 7, and 93 (50.0%) appendicitis patients had a score of 7 or greater ($P = 0.023$). The median Alvarado score in patients with appendicitis was 7 (IQR, 6–8), compared with 6 (IQR, 4–7) in patients without appendicitis ($P = 0.030$). AUC was 0.642 (95% CI, 0.514–0.770).

Modified Alvarado score

Five (22.7%) children without appendicitis had an Alvarado score less than 7, and 55 (33.5%) appendicitis patients had a score of 7 or greater ($P = 0.308$). The score was statistically larger ($P = 0.90$) in patients without appendicitis (median: 5; IQR, 4–6) versus appendicitis patients (median: 6; IQR, 5–7). AUC was 0.611 (95% CI, 0.471–0.751).

Fig. 1

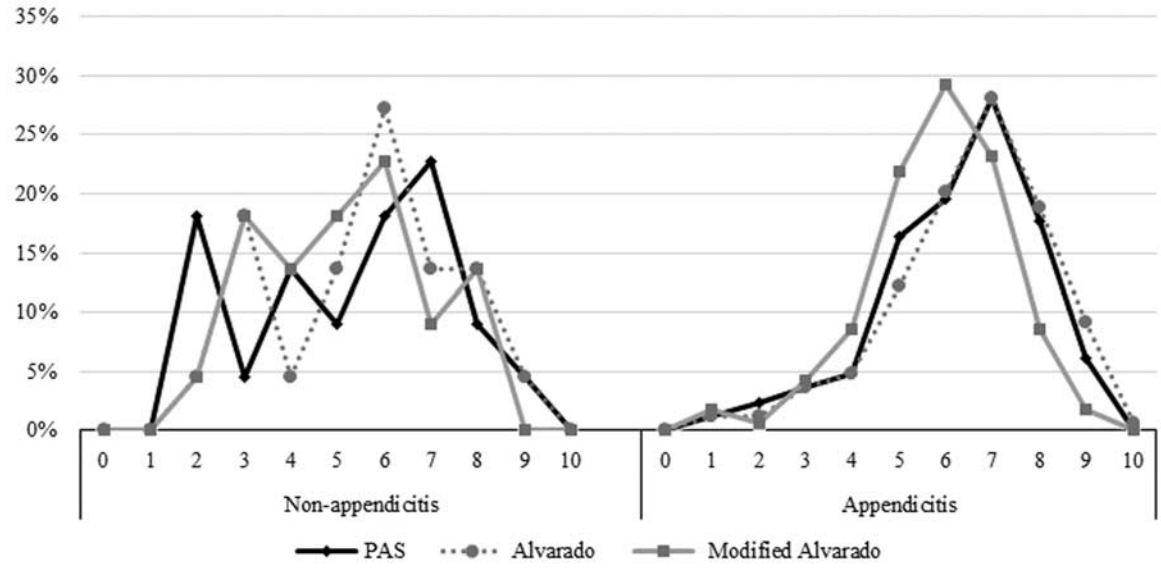


Receiver operator curve of Alvarado score, Modified Alvarado score, and PAS in patients with suspected appendicitis at emergency admission in HINFP (Cartagena, Colombia).

Sensitivity analysis in children above 5 years of age

AUC of the three evaluated scores was very low in children above 5 years of age, and not statistically different ($P = 0.061$). The PAS had an AUC of 0.556 (95% CI, 0.394–0.717); the Alvarado score had an AUC of

Fig. 2



Percentage of patients per score-value in the Alvarado score, Modified Alvarado score, and PAS in patients with and without appendicitis in the sample.

Table 4 Diagnostic utility of the Alvarado score, modified Alvarado score, and Pediatric Appendicitis Score in our sample of pediatric patients, and children above 5 years of age

	Alvarado score		Modified Alvarado score		PAS	
Score value	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
All children						
≥ 1	100.0	0.0	100.0	0.0	100.0	0.0
≥ 2	98.8	0.0	98.2	0.0	98.8	0.0
≥ 3	97.6	4.6	97.6	4.6	96.3	18.2
≥ 4	93.9	22.7	93.3	22.7	92.7	22.7
≥ 5	89.0	27.3	84.8	36.4	87.8	36.4
≥ 6	76.8	40.9	62.8	54.6	71.3	45.5
≥ 7	56.7	68.2	33.5	77.3	51.8	63.6
≥ 8	28.7	81.8	10.4	86.4	23.8	86.4
≥ 9	9.8	95.5	1.8	100.0	6.1	95.5
≥ 10	0.6	100.0	0.0	100.0	0.0	100.0
Children above 5 years of age						
≥ 1	–	–	–	–	–	–
≥ 2	–	–	–	–	100.0	0.0
≥ 3	100.0	0.0	100.0	0.0	98.7	6.7
≥ 4	98.0	13.3	97.3	13.3	96.6	6.7
≥ 5	93.3	13.3	89.3	20.0	93.3	20.0
≥ 6	81.2	20.0	65.8	40.0	76.5	26.7
≥ 7	61.1	60.0	35.6	66.7	55.7	53.3
≥ 8	30.2	73.3	10.7	80.0	24.8	80.0
≥ 9	10.1	93.3	2.0	100.0	6.0	93.3
≥ 10	0.7	100.0	0.0	100.0	0.0	100.0

PAS, Pediatric Appendicitis Score.

0.579 (95% CI, 0.421–0.737), and the modified Alvarado score had an AUC of 0.528 (95% CI, 0.353–0.702) in children above 5 years of age.

Discussion

To our knowledge, this is one of the largest study assessing scores to diagnose appendicitis in a low-income or middle-income pediatric setting [10]. In our developing setting, as other studies have shown, the scores have a relatively poor performance of the overall score.

Several studies have shown evidence that Alvarado score and the PAS [11] do not have an adequate overall accuracy for diagnosing appendicitis. This means that a binary yes/no classification is inadequate for the diagnosis. However, both scores and the modified Alvarado score may be used for risk stratification. This means that stratification in low, intermediate, and high appendicitis risk is likely to be more useful in the clinical setting. A value of 3 or less in the Alvarado and the modified Alvarado score, and a score of 2 or less in the PAS have a sensitivity of 96–97% in diagnosing appendicitis, and a value in any score of 8 or greater has a specificity between 81 and 86% for the three scores. These values, given results from our study, would prove useful in an urban developing setting around the world.

A recent meta-analysis showed in pediatric population with the Alvarado Score a sensitivity of 0.99 (95% CI, 0.83–1.00) for a cutoff value of 5, and a sensitivity of 0.99 (95% CI, 0.83–1.00) for a value of 7. Specificity was 0.81 (95% CI, 0.76–0.85) for that study at a cutoff value of 5, and 0.76 (95% CI, 0.55–0.89) for 7 [10]. These diagnostic values are in disagreement with the results from our study; for a value of 5 or greater, the Alvarado score had a sensitivity of 89% and specificity of 27%.

Children are a special population for the scores currently available to predict appendicitis. Current scores attempt to use clinical symptomatology to make an accurate diagnosis of this disease. However, because of unspecific clinical findings such as age, socioeconomic status, or comorbidities that may affect poor children (i.e. malnutrition), and/or the possibility of intrinsic differences with adult clinical presentation of appendicitis, the PAS was designed by Samuel in 2002 [2] to overcome these shortcomings. The Alvarado and modified Alvarado scoring systems did not seem to be as accurate in diagnosing appendicitis as PAS. The sensitivity of PAS was 100% in their initial validation study, with a specificity of 92% with a cutoff value of 6 or greater.

This study has several limitations. First, the retrospective design could increase information bias. To counteract this, our data collection was carried out by highly trained personnel, and supervised by pediatric specialists who care for patients at HINFP, which use specific protocols. Moreover, selection bias may have been an issue because we retrospectively collected patients without appendectomy [7,10].

In summary, stratification of risk according to prespecified values of the Alvarado score, the modified Alvarado score, and the PAS can be useful in the pediatric clinical setting, if properly performed in the context of evidence-based medicine. No classification of yes/no according to these scores could properly diagnose appendicitis. The group with intermediate risk for appendicitis could benefit with additional diagnostic measures (i.e. ultrasonography or computed tomography scan) in the pediatric population.

Acknowledgements

Conflicts of interest

There are no conflicts of interest.

References

1 Addiss DG, Shaffer N, Fowler BS, Tauxe RV. The epidemiology of appendicitis and appendectomy in the United States. *Am J Epidemiol* 1990; **132**:910–925.
2 Samuel M. Pediatric Appendicitis Score. *J Pediatr Surg* 2002; **37**:877–881.
3 Macklin CP, Radcliffe GS, Merei JM, Stringer MD. A prospective evaluation of the modified Alvarado score for acute appendicitis in children. *Ann R Coll Surg Engl* 1997; **79**:203–205.
4 Kharbanda AB, Taylor GA, Fishman SJ, Bachur RG. A clinical decision rule to identify children at low risk for appendicitis. *Pediatrics* 2005; **116**:709–716.
5 Lintula H, Pesonen E, Kokki H, Vanamo K, Eskelinen M. A diagnostic score for children with suspected appendicitis. *Langenbecks Arch Surg* 2005; **390**:164–170.
6 Van den Broek WT, van der Ende ED, Bijnen AB, Breslau PJ, Gouma DJ. Which children could benefit from additional diagnostic tools in case of suspected appendicitis? *J Pediatr Surg* 2004; **39**:570–574.
7 Kulik DM, Uleryk EM, Maguire JL. Does this child have appendicitis? A systematic review of clinical prediction rules for children with acute abdominal pain. *J Clin Epidemiol* 2013; **66**:95–104.
8 Larrea C, Freire W. Social inequality and child malnutrition in four Andean countries. *Rev Panam Salud Publica* 2002; **11** (5–6):356–364.
9 Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med* 1986; **15**:557–564.
10 Ohle R, O'Reilly F, O'Brien KK, Fahey T, Dimitrov BD. The Alvarado score for predicting acute appendicitis: a systematic review. *BMC Med* 2011; **9**:139.
11 Schneider C, Kharbanda A, Bachur R. Evaluating appendicitis scoring systems using a prospective pediatric cohort. *Ann Emerg Med* 2007; **49**:778–784.