
Prospective Validation of the Pediatric Appendicitis Score in a Canadian Pediatric Emergency Department

Maala Bhatt, MD, MSc, Lawrence Joseph, PhD, Francine M. Ducharme, MD, MSc, Geoffrey Dougherty, MD, MSc, and David McGillivray, MD

Abstract

Objectives: Clinical scoring systems attempt to improve the diagnostic accuracy of pediatric appendicitis. The Pediatric Appendicitis Score (PAS) was the first score created specifically for children and showed excellent performance in the derivation study when administered by pediatric surgeons. The objective was to validate the score in a nonreferred population by emergency physicians (EPs).

Methods: A convenience sample of children, 4–18 years old presenting to a pediatric emergency department (ED) with abdominal pain of less than 3 days' duration and in whom the treating physician suspected appendicitis, was prospectively evaluated. Children who were nonverbal, had a previous appendectomy, or had chronic abdominal pathology were excluded. Score components (right lower quadrant and hop tenderness, anorexia, pyrexia, emesis, pain migration, leukocytosis, and neutrophilia) were collected on standardized forms by EPs who were blinded to the scoring system. Interobserver assessments were completed when possible. Appendicitis was defined as appendectomy with positive histology. Outcomes were ascertained by review of the pathology reports from the surgery specimens for children undergoing surgery and by telephone follow-up for children who were discharged home. Sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV) were calculated. The overall performance of the score was assessed by a receiver operator characteristic (ROC) curve.

Results: Of the enrolled children who met inclusion criteria ($n = 246$), 83 (34%) had pathology-proven appendicitis. Using the single cut-point suggested in the derivation study (PAS 5) resulted in an unacceptably high number of false positives (37.6%). The score's performance improved when two cut-points were used. When children with a PAS of ≤ 4 were discharged home without further investigations, the sensitivity was 97.6% with a NPV of 97.7%. When a PAS of ≥ 8 determined the need for appendectomy, the score's specificity was 95.1% with a PPV of 85.2%. Using this strategy, the negative appendectomy rate would have been 8.8%, the missed appendicitis rate would have been 2.4%, and 41% of imaging investigations would have been avoided.

Conclusions: The PAS is a useful tool in the evaluation of children with possible appendicitis. Scores of ≤ 4 help rule out appendicitis, while scores of ≥ 8 help predict appendicitis. Patients with a PAS of 5–7 may need further radiologic evaluation.

ACADEMIC EMERGENCY MEDICINE 2009; 16:591–596 © 2009 by the Society for Academic Emergency Medicine

Keywords: appendicitis, diagnosis, accuracy, pediatrics, clinical score

From the Division of Emergency Medicine (MB, DM), Division of General Pediatrics (GD), Department of Pediatrics, Montreal Children's Hospital, McGill University, Montreal, Quebec; the Division of Clinical Epidemiology, Royal Victoria Hospital, McGill University (LJ), Montreal, Quebec; and the Department of Pediatrics, Hôpital Sainte Justine, University of Montreal (FMD), Montreal, Quebec, Canada.

Received January 15, 2009; revision received March 17, 2009; accepted April 6, 2009.

Presented at the Pediatric Academic Societies Meeting, San Francisco, CA, April 29–May 2, 2006.

This work has been supported by a grant from the Montreal Children's Hospital Research Institute.

Address for correspondence and reprints: Maala Bhatt, MD; E-mail: maala.bhatt@muhc.mcgill.ca.

Appendicitis is the most common atraumatic surgical abdominal disorder in children over 2 years of age,^{1–3} with approximately 70,000 appendectomies performed each year in the United States. The diagnosis of appendicitis is problematic in children because many present with signs and symptoms that mimic other common, but self-limited, causes of abdominal pain. It is estimated that one-third of children with appendicitis have been previously evaluated by a physician for their symptoms, resulting in initial misdiagnosis rates from 28% to 57% in school-aged children.⁴ The clinical challenge is to diagnose appendicitis early enough to prevent progression to perforation, while minimizing the number of negative

appendectomies that are performed. Ultrasound (US) and computed tomography (CT) are commonly used as diagnostic aids in pediatric appendicitis; however, both have important limitations. The accuracy of ultrasound is known to be operator dependent and is affected by the presence of certain patient characteristics (pain, obesity) and by a lack of confidence in a negative result due to the difficulty in visualizing a noninflamed appendix.⁵⁻⁹ CT is a highly accurate diagnostic tool; however, there is evidence that exposure to ionizing radiation in childhood likely increases lifetime mortality risk from cancer.^{10,11}

Clinical scoring systems have been investigated as alternatives or adjuncts to diagnostic imaging.¹²⁻¹⁸ They are safe, inexpensive, time-efficient tools that have the potential to improve patient outcomes. The first appendicitis score specific to children was published by Samuel in 2002.¹⁹ In a population of children who were referred to, and assessed by, surgeons, Samuel provided preliminary evidence that the Pediatric Appendicitis Score (PAS) might be able to accurately distinguish those children with and without appendicitis, using a single cut-point (sensitivity 100% [95% confidence interval {CI} = 99.2% to 100%], specificity 92% [95% CI = 89.0% to 94.2%]). A recent prospective validation study, using a cohort of patients from a pediatric emergency department (ED), failed to reproduce the accuracy shown in the derivation study.²⁰ Applying the PAS as suggested by Samuel in this study would have resulted in a 12% missed appendicitis rate and a 45% negative appendectomy rate. This study only investigated the cut-point described by Samuel and did not explore whether other cut-points may have improved the performance of the score as a diagnostic strategy. A second study applied the PAS to a cohort of children; however, the authors altered the scoring system by omitting two score elements on an unspecified number of patients who were included in their final analysis and did not assess the cut-points as proposed by Samuel.²¹

The primary objective of this study was to assess the performance of the PAS in a cohort of children who were presenting to a pediatric ED with abdominal pain suggestive of appendicitis. Specifically, we sought to determine the diagnostic properties of the optimal cut-point defined by Samuel for diagnosing appendicitis and whether other cut-points could be used to optimize decision-making in our population and clinical setting. We also wanted to assess the potential impact of the PAS on patient outcomes (negative appendectomy rate, missed appendectomy rate) and estimate the reduction of imaging investigations by retrospectively applying the cut-points suggested by our data to our population.

METHODS

Study Design

A prospective observational study was conducted in the ED of the Montreal Children's Hospital between November 2003 and July 2005. The study received approval from the research ethics board at the Montreal Children's Hospital. Informed written consent was

obtained from all parents or legal guardians, and assent was obtained from children 7 years or older.

Study Setting and Population

This urban, tertiary care pediatric teaching hospital has an ED census of 65,000 visits per year. The hospital serves a population of 3 million in the greater Montreal area and is a designated referral center for approximately 23% of the population of Quebec.

Children between the ages of 4 and 18 years with less than 3 days of abdominal pain, and in whom the emergency physician (EP) considered a diagnosis of appendicitis, were eligible for enrollment in the study. Patients could have been self-referred or physician-referred to the ED for their abdominal pain. Physicians consider a diagnosis of appendicitis based on the assimilation of information obtained from their clinical examination and results from additional testing. At our institution, it is usual practice to obtain a complete blood count (CBC) on all children with suspected appendicitis. Children were excluded if they were non-verbal, had a previous appendectomy, or had chronic abdominal pathology (e.g., inflammatory bowel disease, a history of complex abdominal surgery, or significant congenital abdominal anomalies) that may have interfered with the assessment of the abdomen. Patients were approached for participation in the study either by a research assistant or by the EP.

Study Protocol

After obtaining informed consent, the EP completed a one-page data collection form. The form contained information about patient age, sex, date and time of the examination, the date and time of the onset of symptoms, and each of the eight PAS components (Table 1). All data collection forms were completed prior to obtaining any imaging investigations or surgical consultation. All physicians in medical (nonsurgical) specialties at the second-year resident level or higher could enroll patients. This level of training was chosen because it was felt that after 2 years of postgraduate training (equivalent to a family medicine residency in Canada) physicians should have the clinical skills necessary to make an accurate physical assessment. Pediatric emergency medicine attending physicians and fellows

Table 1
PAS¹⁹

Diagnostic Indicators	Score Value
Cough <i>or</i> percussion <i>or</i> hop tenderness	2
Anorexia	1
Pyrexia	1
Nausea/emesis	1
Tenderness in RLO	2
Leukocytosis > 10,000	1
Polymorphonuclear neutrophilia	1
Migration of pain	1
<i>Total</i>	10

PAS = Pediatric Appendicitis Score; RLO = right lower quadrant.

were introduced to the data collection form and study definitions prior to the start of the study. Residents from other specialties working in the ED for shorter periods received this same introduction during an orientation session on the first day of their rotation. Enrolling physicians were not informed of the weighting of elements used in the scoring system, and there was no indication of score values on the data collection form. Decisions for laboratory or imaging investigations, surgical consultation, and disposition from the ED were left to the discretion of the treating physician. At our institution, when a physician is unsure of the etiology of abdominal pain but suspects appendicitis, he or she typically orders imaging investigations and/or a surgical consultation to help refine the diagnosis. Disposition from the ED is made at the physician's discretion after all investigations have been obtained.

When possible, a second physician performed an independent assessment of the child and completed an identical data collection form to test the interobserver reliability of the score. The second assessor was identified by the enrolling physician and was often the attending staff or resident working with the enrolling physician or another attending physician working in the ED. It was recommended that this second evaluation be done immediately following the first; however, the exact times of the evaluations were not documented.

Patients who were discharged directly home from the ED were contacted by telephone at 1 month to verify final outcome. Although most children with abdominal pain caused by appendicitis will progress to perforation within 72 hours of the onset of symptoms, a delay of 1 month was chosen to ascertain outcome for study logistics. Patients or parents were asked if they or their child had an appendectomy at the Montreal Children's Hospital or elsewhere since their ED visit. If a patient underwent an appendectomy at the study site or elsewhere, the medical record was obtained and the pathology was reviewed. Appendicitis was defined as appendectomy with positive histology. A negative appendectomy was defined as an appendectomy with negative histology. Missed appendicitis was defined as a child who was discharged home from the ED but within 1 week had an appendectomy with positive histology.

For analysis, the patients were separated into two groups: those with histology-confirmed appendicitis and those without appendicitis. The latter group included children who underwent appendectomy but who had negative histology. The number and results of imaging investigations with CT or US and the results from the pathology reports were abstracted from the medical records. The components of the PAS were used exactly as described by Samuel. However, as he did not provide definitions for polymorphonuclear neutrophilia or pyrexia, we defined polymorphonuclear neutrophilia as $\geq 75\%$ neutrophils on CBC and pyrexia as $>38^\circ\text{C}$ (oral or rectal). When a patient had two scores by two physicians for interobserver reliability testing, the score from the first physician was always used in the primary analysis to calculate the PAS.

Data Analysis

Data were entered by one author (MB) on a monthly basis into a Microsoft Access (Microsoft Inc., Redmond, WA) database. For each patient, information from the data collection form, final outcome (discharged home, discharged but returned and had appendectomy, or appendectomy performed at initial visit), and results of the surgical pathology for patients who had appendectomies were entered. The PAS was calculated as detailed by Samuel.¹⁹ The sensitivity and specificity, with 95% CI, were calculated for each value of the score (1–10). In addition, positive predictive value (PPV) and negative predictive value (NPV) were calculated for the optimal cut-points. The potential negative appendectomy rate was calculated as the number of false positives divided by the number of patients taken to the operating room for an appendectomy (this calculation assumes that all patients with appendicitis were taken to the operating room for an appendectomy). The potential missed appendicitis rate was calculated as the number of false negatives divided by the number of patients with appendicitis. The difference between means with 95% CIs were calculated for all baseline characteristics, separated by group. A receiver operator characteristic (ROC) curve was created to assess the overall performance of the score. Cohen's kappa coefficient was calculated to measure interobserver agreement on the subset of patients evaluated by two physicians. A reliable estimate was considered to have a kappa value of higher than 0.6. Data were analyzed using SAS statistical software (Version 9.1, SAS Institute Inc., Cary, NC).

RESULTS

Characteristics of Study Subjects

A convenience sample of 275 patients was enrolled between November 2003 and July 2005. Twenty-nine patients were excluded for failure to meet inclusion criteria (14 patients with abdominal pain for greater than 72 hours or age less than 4 years) or for not having a CBC performed (15 patients). Of the remaining 246 patients, the mean age was 10.9 years (standard deviation [SD] ± 3.4 years) and no patients had missing data for any of the score components. Ninety-five children were taken to the operating room for appendectomies. Of these children, 83 (34% of the total population) had pathology-proven appendicitis, including 14 (16.9%) who had perforated appendicitis. Twelve (12.6%) had negative appendectomies. All patients were contacted by telephone at 1 month to verify final outcome. There were no cases of missed appendicitis. The patients with and without appendicitis were similar, with the exception of mean PAS, and their characteristics are described in Table 2. The distribution of scores is shown in Figure 1.

Main Results

An ROC curve (Figure 2) was constructed to assess PAS performance in our population and yielded an area under the curve of 0.895. The best cut-point, calculated to maximize the sensitivity and specificity and found at

Table 2
Study Subject Characteristics

Patient Characteristics	Appendicitis (n = 83)	No Appendicitis (n = 163)	Difference in Means (95% CI)
Mean age, yr (±SD)	11.6 (±3.8)	10.5 (±3.1)	1.1 (0.4 to 2.3)
Sex (female)	35%	43%	8% (-5 to 20)
Mean symptom duration, days	1.5	1.6	-0.1 (-0.3 to 0.1)
Mean PAS (±SD)	7.5 (±1.2)	4.3 (±1.5)	3.1 (2.6 to 3.6)
Imaging investigations	37 (45%)	54 (33%)	12% (-1 to 24)

PAS = Pediatric Appendicitis Score.

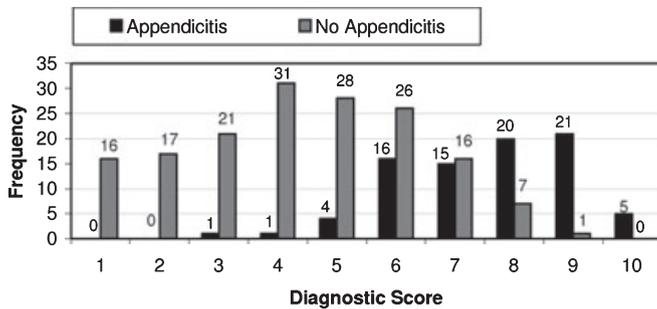


Figure 1. Distribution of scores in patients with and without appendicitis.

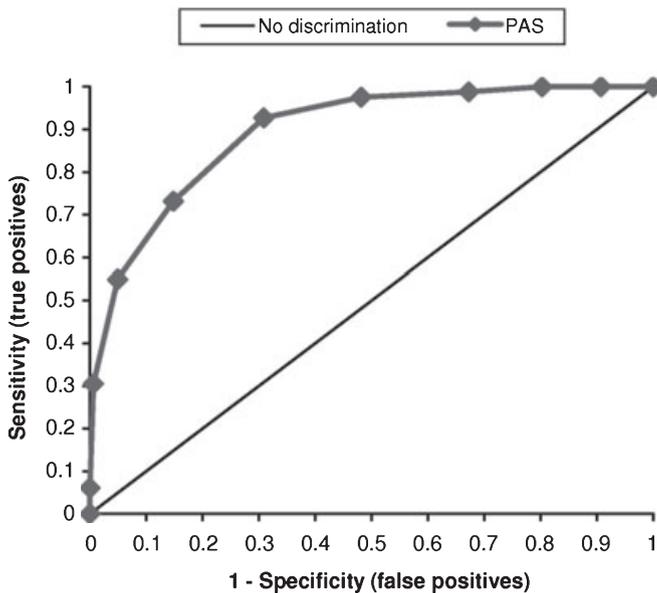


Figure 2. Receiver operating characteristic curve.

the upper left-hand corner on the ROC curve, was the same as Samuel’s, where children with scores of 6 or more are taken to the operating room for an appendectomy and children with scores of 5 or less are discharged home with a diagnosis of “no appendicitis.” At this point, the score was very sensitive (92.8% [95% CI = 85.1% to 96.6%]) but not very specific (69.3% [95% CI = 61.9% to 75.9%]). Using this cut-point, there would have been 50 (37.6%) negative appendectomies, and six (7.2%) cases of missed appendicitis. The sensitivity,

specificity, potential negative appendectomy rate, and missed appendicitis rate were calculated for each score value (Table 3). No other single cut-point offered an improved sensitivity and specificity.

Interobserver scores were obtained in 37 (14.6%) of the 246 patients. The kappa coefficient was 0.65 (95% CI = 0.48 to 0.81), indicating substantial agreement beyond chance between the raters. Forty-six percent of the pairs had perfect agreement, and 92% (n = 34) agreed within two points. The management of five patients would have changed on the basis of the interobserver scores. Four of these patients would have undergone further imaging instead of being discharged home directly. None of these four patients had appendicitis. The remaining patient would have been taken directly to the operating room instead of undergoing further imaging. This patient had pathology-confirmed appendicitis.

DISCUSSION

In this prospective validation study of the PAS using a convenience sample of children aged 4 to 18 years presenting with abdominal pain suggestive of appendicitis, we were unable to reproduce the accuracy of the PAS reported by Samuel in his derivation study. Despite a good estimate of overall accuracy by the ROC curve (area under the curve = 0.895) in our population, using the single cut-point proposed by Samuel to decide if patients should be discharged home (PAS ≤ 5) or undergo an appendectomy (PAS ≥ 6) would have resulted in an unacceptably high rate of negative appendectomies (37.6%).

However, the score’s performance did improve when two thresholds were used to direct patient care: one to decide who could be discharged home and one to decide who should have an appendectomy. If a score of 4 or less was used to discharge patients home without further investigation, two patients (2.4%) with appendicitis would have been erroneously discharged home. At this cut-point, the score had a sensitivity of 97.6% (95% CI = 91.6% to 99.3%) with an NPV of 97.7% (95% CI = 92.0% to 99.4%). If a score of 8 or more was used to determine the need for appendectomy, eight children (8.8%) would have undergone a negative appendectomy. At this cut-point, the score had a specificity of 95.1% (95% CI = 90.6% to 97.5%) and a PPV of 85.2% (95% CI = 73.4% to 92.3%). Patients with scores of 5, 6, and 7 have an uncertain diagnosis, with the score unable to adequately distinguish between those who do or do not have appendicitis. Children with scores in this

Table 3
Score Performance at Each Cut-point (Test Positive if Score \geq Cut-point)

PAS	Sensitivity, % (95% CI)	Specificity, % (95% CI)	% Negative Appendectomy	% Missed Appendicitis
1	100 (95.6–100)	4.3 (2.1–8.6)	65.3	0
2	100 (95.6–100)	9.0 (8.5–9.4)	63.9	0
3	100 (95.6–100)	20.2 (14.8–27.0)	61.0	0
4	98.8 (93.5–99.8)	33.1 (26.4–40.7)	57.8	1.2
5	97.6 (91.6–99.3)	52.1 (44.5–59.7)	48.4	2.4
6	92.8 (85.1–96.6)	69.3 (61.9–75.9)	37.6	7.2
7	73.5 (63.1–81.8)	85.3 (79.0–89.9)	22.4	26.5
8	55.4 (44.7–65.6)	95.1 (90.6–97.5)	8.8	44.6
9	31.3 (22.4–42.0)	99.4 (96.6–99.9)	1.2	68.7
10	6.0 (2.6–13.3)	100 (97.7–100)	0	94.0

PAS = Pediatric Appendicitis Score.

range require further imaging investigations to determine the etiology of their abdominal pain. If this decision-making process had been applied to our population, 37 (41%) imaging investigations would have been avoided because patients would have either been discharged home with a low likelihood of appendicitis or taken directly to the operating room for a high suspicion of appendicitis without any further investigations. Using this management strategy, the negative appendectomy rate would have decreased from 12.6% to 8.8%; however, the missed appendicitis rate would have increased from 0% to 2.4%. These results are based on the assumption that children with scores of 5, 6, and 7 receive an accurate diagnosis with further imaging investigations. Minimizing the number of imaging investigations is particularly important in the pediatric population for two reasons. First, when using US to diagnose appendicitis, definitive management decisions may be delayed because the results of the study may not help with the diagnosis. Second, a single CT scan has the potential to harm a patient by increasing his or her lifetime mortality risk from cancer.¹¹ Although we have not investigated this issue, it is conceivable that the utility of the PAS may be heightened in settings that do not have the advanced pediatric expertise found in children's hospitals. By standardizing the collection, assimilation, and interpretation of patient information, the PAS has the potential to have a positive effect on patient outcomes while minimizing the number of imaging investigations ordered.

To our knowledge, our study is the first to examine the interobserver reliability of the score, which showed good agreement. Two previous studies have investigated the use of the PAS in the ED. The first, by Schneider and colleagues²⁰, prospectively validated the PAS in a convenience sample of 588 patients who had a surgical consultation for appendicitis. Their population, with a mean age of 11.9 years and prevalence of appendicitis of 34%, was similar to ours, and they also observed similar results when using a cut-point of 5, as suggested by Samuel. The authors could not support the use of Samuel's strategy, as this would have resulted in 36 (12%) cases of missed appendicitis and 136 (45%) patients with a negative appendectomy. Unfortunately, the authors did not explore whether other cut-points might have improved the performance

of the PAS in their population. A second study by Goldman and colleagues²¹ explored the utility of alternative PAS cut-points in a convenience sample of 849 children with a chief complaint of abdominal pain for less than 7 days duration.²¹ Using cut-points of PAS \leq 2 and PAS \geq 7, the score had a sensitivity of 97.6% and specificity of 96.0%. However, and critically, the authors modified the PAS in an unspecified number of children by including patients who did not have a CBC performed at the time of their ED evaluation. In these patients, they ignored the missing score elements and simply summed the remaining components. This subset of patients with missing data was included for analysis with patients who had complete data for all score components. The potential loss of three PAS points with the omission of the CBC would have resulted in the misclassification of patients in all categories. Given this methodologic flaw, we are unable to accurately or reliably compare our study results to theirs.

LIMITATIONS

First, we did not track missed but eligible patients. We used a convenience sample of children who were enrolled in the study at the physician's discretion. Although physicians were encouraged to enroll every child in whom they were considering a diagnosis of appendicitis, due to the busy work environment in the ED, some patients may have been missed. There may have been an overrepresentation of equivocal cases in our sample if physicians tended to enroll children when the diagnosis was uncertain as opposed to when they were more certain of disposition (home or operating room). This could have contributed to the discrepancy between our results and Samuel's results, as his cohort contained patients with a higher prevalence of disease (63%) compared to ours (34%). Second, we allowed all medical physicians at the second year resident level or higher to enroll patients and were unable to determine the effect of training level on the performance of the score, due to prohibitively large amounts of missing data for training level.

In addition to the above, several other factors may help explain why the PAS did not perform as well in our population as reported in Samuel's original article. First, the definitions we used for pyrexia and neutrophilia may

have differed from Samuel's as these terms were not defined in his article. Second, our population included children presenting to the ED with abdominal pain suggestive of appendicitis, not just those who had been prescreened by a physician and referred to a surgeon for further evaluation, as in Samuel's study. Finally, a factor related to the derivation of the rule may also have contributed to our discrepant results. The PAS was created using stepwise regression to guide the inclusion of variables into the model. Failure of clinical prediction rules to achieve comparable performance in future validation studies is a well-recognized problem in models developed using these methods.²² Using stepwise rules based on p-values to guide inclusion or exclusion of independent variables in a model tends to lead to overfitting to the data, ignoring of model uncertainty, and ultimately, to poor generalizability.

Using the hierarchy of evidence (Level 4 to Level 1) for clinical decision rules described by McGinn et al.,²³ the PAS used as suggested by Samuel has not progressed past the most basic level of evidence (Level 4). Schneider's group and our group attempted to validate the PAS in the hands of medical physicians and in a less selective population of children. This failure to show equivalent or acceptable performance in a different population of children and in the hands of medical physicians instead of surgeons has not allowed it to progress further in the hierarchy of evidence.

CONCLUSIONS

An accurate, rigorously developed scoring system for childhood appendicitis would be extremely valuable to clinicians. We believe that it is unrealistic to expect that a score with a single cut-point will be able to accurately categorize all or almost all patients with and without appendicitis, given the extremely varied presentation of childhood appendicitis. However, using a score to decide who should be discharged home directly, who should receive imaging because of equivocal findings, and who should be taken directly for an appendectomy seems more realistic and almost as useful. Although we cannot recommend the immediate adoption of our proposed cut-points into clinical practice without further investigation and validation, we believe that this strategy shows promise and should be prospectively evaluated in a subsequent cohort of children.

References

1. Reynolds SL, Jaffe DM. Children with abdominal pain: evaluation in the pediatric emergency department. *Pediatr Emerg Care.* 1990; 6:8-12.
2. Reynolds S L, Jaffe DM. Diagnosing abdominal pain in a pediatric emergency department. *Pediatr Emerg Care.* 1992; 8:126-8.
3. Scholer SJ, Pituch K, Orr DP, Dittus RS. Clinical outcomes of children with acute abdominal pain. *Pediatrics.* 1996; 98:680-5.
4. Rothrock SG, Pagane J. Acute appendicitis in children: emergency department diagnosis and management. *Ann Emerg Med.* 2000; 36:39-51.
5. Hormann M, Scharitzer M, Stadler A, Pokieser P, Puig S, Helbich T. Ultrasound of the appendix in children: is the child too obese? *Euro Radiol.* 2003; 13:1428-31.
6. Ramachandran P, Sivit CJ, Newman KD, Schwartz MZ. Ultrasonography as an adjunct in the diagnosis of acute appendicitis: a four year experience. *J Pediatr Surg.* 1996; 31:167-9.
7. Sivit CJ, Newman KD, Boenning DA, et al. Appendicitis: usefulness of US in a pediatric population. *Radiology.* 2002; 185:549-52.
8. Crady SK, Jones JS, Wyn T, Luttenton CR. Clinical validity of ultrasound in children with suspected appendicitis. *Ann Emerg Med.* 1993; 22:1125-9.
9. Quillin SP, Seigel MJ. Appendicitis: efficacy of color Doppler sonography. *Radiology.* 1994; 191:557-60.
10. Doria A, Moineddin R, Kellenberger CJ, et al. US or CT for diagnosis of appendicitis in children and adults? A meta-analysis. *Radiology.* 2006; 241:83-94.
11. Brenner DJ, Elliston CD, Hall EF, Berdon W. Estimated risks of radiation-induced fatal cancer from pediatric CT. *Am J Roentgen.* 2001; 176:289-96.
12. Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med.* 1986; 15:557-64.
13. Bond GR, Tully SB, Chan LS, Bradley RL. Use of the MANTRELS score in childhood appendicitis: a prospective study of 187 children with abdominal pain. *Ann Emerg Med.* 1990; 19:1014-8.
14. Kalan M, Talbot D, Cunliffe WJ, Rich AJ. Evaluation of the modified Alvarado score in the diagnosis of acute appendicitis: a prospective study. *Ann R Coll Surg Engl.* 1994; 76:418-9.
15. 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; 79:203-5.
16. Ohmann C, Franke C, Yang O. Diagnostic scores for acute appendicitis: Abdominal Pain Study Group. *Euro J Surg.* 1995; 161:273-81.
17. Tzanakis NE, Efstathiou SP, Danulidis K, et al. A new approach to accurate diagnosis of acute appendicitis. *World J Surg.* 2005; 29:1151-6.
18. van den Broek WT, Bijnen BB, Rijbroek B, Gouma DJ. Scoring and diagnostic laparoscopy for suspected appendicitis. *Euro J Surg.* 2002; 168:349-54.
19. Samuel M. Pediatric appendicitis score. *J Pediatr Surg.* 2002; 37:877-81.
20. Schneider C, Kharbanda A, Bachur R. Evaluating appendicitis scoring systems using a prospective pediatric cohort. *Ann Emerg Med.* 2007; 49:778-84.
21. Goldman RD, Carter S, Stephens D, Antoon R, Mounstephen W, Langer JC. Prospective validation of the pediatric appendicitis score. *J Pediatr.* 2008; 153:278-82.
22. Raftery AE. Bayesian model selection in social research. *Sociological Method.* 1995; 25:111-63.
23. McGinn TG, Guyatt GH, Wyer PC. Users Guide to the Medical Literature: XXII How to use articles about clinical decision rules. *JAMA.* 2000; 284:79-84.