

# Prospective Validation of the Pediatric Appendicitis Score

RAN D. GOLDMAN, MD, SUSAN CARTER, BSc, DEREK STEPHENS, MD, ROULA ANTOON, MD, WILLIAM MOUNSTEPHEN, MD, AND JACOB C. LANGER, MD

**Objective** To prospectively validate the Pediatric Appendicitis Score (PAS), developed on a cohort of children with abdominal pain suggestive of appendicitis, in unselected children with abdominal pain who present to the emergency department.

**Study design** Over a 19-month period, we prospectively recruited children 1 to 17 years old who came to our tertiary pediatric emergency department, with a chief complaint of abdominal pain of duration less than 7 days. PAS components included fever  $>38^{\circ}$  C, anorexia, nausea/vomiting, cough/percussion/hopping tenderness (2 points), right-lower-quadrant tenderness (2 points), migration of pain, leukocytosis  $>10\,000$  cells/mm<sup>3</sup>, and polymorphonuclear neutrophilia  $>7500$  cells/mm<sup>3</sup>. A follow-up call was made to verify final outcome. Sensitivity, specificity, and the receiver operating characteristic curve of the PAS with respect to diagnosis of appendicitis were calculated.

**Results** We collected data on 849 children. 123 (14.5%) had pathologic study-proven appendicitis. Mean (median, range) score for children with appendicitis and without appendicitis was 7.0 (7, 2-10) and 1.9 (1, 0-9), respectively. If a cutoff PAS of  $\leq 2$  was used to discharge patients without further investigation, only 3 (2.4%) with appendicitis would be sent home. If a PAS of  $\geq 7$  was used to take children to the operating room without further investigation, only 29 (4%) would not have appendicitis. For the PAS the area under the receiver operator curve was 0.95.

**Conclusions** The PAS is useful, because a value  $\leq 2$  (found in 73% of children without appendicitis) has high validity for ruling out appendicitis, and a score  $\geq 7$  (found in 61% of children with appendicitis) has a high validity for predicting the presence of appendicitis. Children with PAS of 3 to 6 (37% with appendicitis and 23% without appendicitis in this study) should undergo further investigation such as observation, ultrasonography, or computed tomography. (*J Pediatr* 2008;153:278-82)

Abdominal pain is one of the most common complaints in the pediatric population, and acute abdominal pain is the most common surgical emergency in childhood.<sup>1</sup> Early, accurate diagnosis potentially could decrease perforation rate and its complications.<sup>2</sup> Early diagnosis of “no appendicitis” or “appendicitis” on the basis of a pediatric appendicitis score potentially could decrease emergency department time and resource use and could avoid time, cost, and risks for further evaluation.<sup>3-6</sup>

Failure to diagnose appendicitis is caused by poor discrimination between the presentations of acute abdominal pain caused by appendicitis and other urgent conditions.<sup>7</sup> Although missing the diagnosis of appendicitis can result in significant morbidity and death, overdiagnosis exposes children to unnecessary surgical procedures in up to 46% of the cases,<sup>6,8-10</sup> with potential unnecessary exposure to radiation because of computed tomography (CT) scanning and postsurgical complications.<sup>11</sup> For this reason, many authors have suggested that a scoring system be developed that would incorporate the clinical and laboratory findings seen in patients with appendicitis.

Several scoring systems have been reported previously in the literature,<sup>12-24</sup> most in the adult population. However, retrospective design,<sup>12,14,22</sup> lack of clinical evaluation,<sup>13</sup> addition of other diagnostic modalities,<sup>16</sup> and small sample size<sup>18</sup> made the creation of a diagnostic clinical score challenging. Inherent differences between children and adults with respect to their ability to communicate subjective symptoms and the complexity in assessment of pain in children are two of the challenges in identifying pediatric appendicitis early in the emergency department.

From the Pediatric Research in Emergency Therapeutics (PRETx) Program, Division of Pediatric Emergency Medicine, BC Children's Hospital, Department of Pediatrics, University of British Columbia, and Child & Family Research Institute, Vancouver, British Columbia (R.G.), the Division of Pediatric Emergency Medicine, (S.C., R.A., W.M.), the Population Health Sciences, Research Institute, (D.S.), and the Division of General Surgery, Department of Surgery (J.L.), The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada.

Submitted for publication Oct 15, 2005; last revision received Oct 5, 2007; accepted Jan 24, 2008.

Reprint requests: Ran D. Goldman, MD, Division of Pediatric Emergency Medicine, BC Children's Hospital, Department of Pediatrics, University of British Columbia, 4480 Oak St, Vancouver, BC V6H 3V4, Canada. E-mail: rgoldman@cw.bc.ca.

0022-3476/\$ - see front matter

Copyright © 2008 Mosby Inc. All rights reserved.

10.1016/j.jpeds.2008.01.033

---

CT	Computed tomography	ROC	Receiver operator characteristic
PAS	Pediatric Appendicitis Score		

---

Recently, Samuel from England published a simple pediatric appendicitis score (PAS) on the basis of a cohort of children 4 to 15 years old.<sup>24</sup> The PAS ranges from 0 to 10, and it includes components from history: migration of pain (1 point), anorexia (1 point), nausea/vomiting (1 point); physical examination: fever  $>38^{\circ}\text{C}$  (1 point), cough/percussion/hopping tenderness (2 points), right-lower-quadrant tenderness (2 points); and laboratory results: leukocytosis  $>10,000$  cells/mm<sup>3</sup> (1 point) and polymorphonuclear neutrophilia  $>7500$  cells/mm<sup>3</sup> (1 point). When tested on the same population, the sensitivity was 100%, specificity was 92%, positive predictive value was 96%, and negative predictive value was 99%. A score  $\geq 6$  was highly associated with appendicitis. The aim of this study was to prospectively validate the PAS score in a large group of pediatric patients presenting with abdominal pain to the emergency department.

## METHODS

The study was approved by the Hospital for Sick Children Research Ethics Board. Between September 1, 2003, and March 31, 2005, we prospectively recruited children 1 to 17 years old who came to our tertiary urban pediatric emergency department in Toronto, Ontario, Canada, with a chief complaint of abdominal pain of less than 7 days duration. Our emergency department census is 50,000 children a year, with 24-hour coverage by a pediatric emergency medicine staff and an in-hospital pediatric surgery fellow.

We excluded children who had a known diagnosis of appendicitis by ultrasonography or CT on arrival to the emergency department, children with abdominal pain for greater than 7 days, and patients who had a history of prior appendectomy.

After obtaining informed consent from the parents and an informed assent from children 7 years or older, the staff physicians were asked to complete a structured form that included history of current illness, chronic illness, and findings on physical abdominal examination. One research nurse collected information on laboratory test results and disposition. If blood tests were not obtained because of clinical appearance as decided by treating faculty physician, a score was calculated on the basis of components of history and physical examination.

For patients who were admitted to the hospital, the research nurse reviewed the in-hospital chart, the surgical report after an operation, and the pathology report if an appendectomy was performed. After 5 to 7 days, the research nurse conducted a follow-up call with the family to document final diagnosis after discharge from the emergency department or the hospital (appendicitis or not, respectively), and to learn whether the family sought further medical consultation. Children were identified as a study group (appendicitis) or control group (no-appendicitis) only after the follow-up interview. If the children still had abdominal complaints at the time of the follow-up call, they were enrolled to the control group, because we anticipated appendicitis to be diagnosed within 5 to 7 days of onset of signs and symptoms.

Components of the PAS were entered on the Microsoft Excel program (Microsoft Corporation, Redmond, Wash) in a double data entry system. PAS was calculated for all patients. Descriptive statistics were done with SAS version 9.1 (SPSS Inc, Chicago, Illinois) to compare the study group of children with appendicitis to the control group of children without appendicitis. The  $\chi^2$  test was used for comparison of frequencies and Student's *t* tests or analysis of variance for comparisons of continuous variables between the study group and the control group. Calculation of correlations was done with the Pearson test or Spearman test, whenever appropriate. A receiver operating characteristic (ROC) curve was constructed to assess sensitivity and specificity and optimal cut points for the PAS to diagnose appendicitis. Area under the curve and corresponding 95% CI were calculated. Finally, we used the Microsoft Excel program to plot a graph describing association between PAS and the rate of patients in both study groups.

The sample size calculation was based on a 2-sided *z*-test at a significance level of 0.05. We needed a sample of 103 children from the acute appendicitis (study group) group and 103 from the non-appendicitis (control) group to achieve 80% power to detect a difference of 0.07 between the area under the ROC curve under our null hypothesis of 0.85 and an AUC under the alternative hypothesis of 0.92.

## RESULTS

A total of 1060 patients were approached to participate in the study, and 849 (80%) children 1 to 17 years old were recruited to the study and fulfilled the inclusion and exclusion criteria. We had no study attrition; all patients for whom consent was obtained were included in the study. A total of 123 patients (14.5%) had pathology-proven appendicitis. Comparison between the study group and the control group are presented in Table I. The mean (median, SD, range) score for children with appendicitis and without appendicitis was 7 (7, 2.2, 2-10) and 1.9 (1, 1.9, 0-9), respectively. If the threshold for making a diagnosis of "not appendicitis" had been a PAS of  $\leq 2$ , only 3 (2.4%) patients with appendicitis would have been sent home from the emergency department. If the threshold for making a diagnosis of "appendicitis" had been a PAS of  $\geq 7$ , only 29 (4%) patients without appendicitis would have undergone surgery.

The sensitivity, specificity, and cumulative area under the ROC curve with 95% CI for the different PAS cutoff points are presented in Table II. The total area under the ROC curve was 0.948 and the curve is presented in Figure 1. The ROC curve shows that the score is highly sensitive to diagnosis of appendicitis because the area under the curve is higher with increment of the score. Figure 2 presents the rate of patients at each PAS cut-off point.

## DISCUSSION

We found PAS score<sup>24</sup> to be valid for the diagnosis of appendicitis when the PAS was 7 or greater, and for the exclusion of appendicitis when the PAS was 2 or under. For

**Table I. PAS in the study cohort of children with and without appendicitis**

Variable	Appendicitis study group (n = 123)	No-appendicitis control group (n = 726)	P value	95% CI for OR
Mean PAS (SD)	7.0 (2.2)	1.9 (1.9)	<.001	CI for diff. of means: (4.7, 5.5)
Median PAS	7	1		
History				
Migration of pain	57 (46%)	70 (10%)	<.001	8.1 (5.3, 12.5)
Nausea or vomiting	92 (75%)	335 (46%)	<.001	3.5 (2.2, 5.3)
Anorexia	84 (68%)	258 (36%)	<.001	3.9 (2.6, 5.9)
Physical examination				
Fever (>38° C)	73 (59%)	154 (21%)	<.001	5.4 (3.6, 8.1)
Cough/percussion/hopping tenderness in the right lower quadrant	88 (72%)	68 (9%)	<.001	24.3 (15.3, 38.7)
Tenderness over the right iliac fossa	98 (80%)	122 (17%)	<.001	19.4 (12.0, 31.4)
Laboratory results				
Leukocytosis WBC ≥ 10 000/mm <sup>3</sup>	98 (88%) (n=112)	84 (43%) (n=196)	<.001	30.0 (18.2, 49.1)
Polymorphonuclear neutrophilia ≥ 7500/mm <sup>3</sup>	84 (84%) (n=101)	67 (36%) (n=188)	<.001	21.2 (13.4, 33.4)

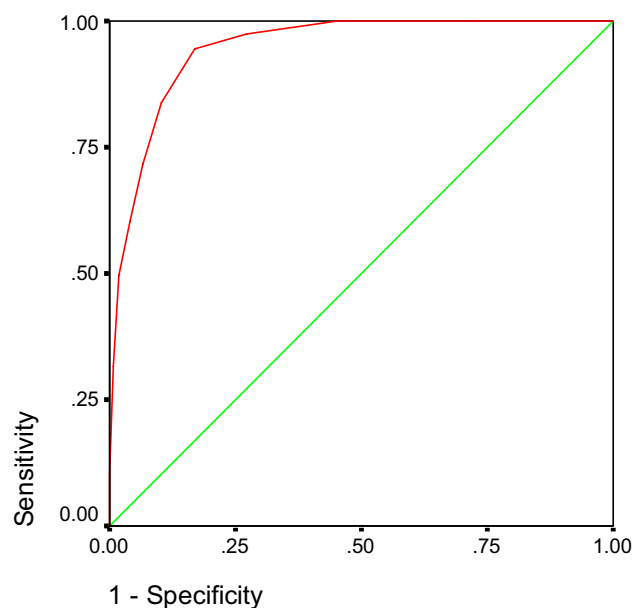
CI, Confidence interval; OR, odds ratio; WBC, white blood cell count.

**Table II. Sensitivity and specificity for PAS**

PAS cutoff point	Sensitivity	Specificity
≥0	1.00	0.27
≥1	0.87	0.55
≥2	0.68	0.73
≥3	0.50	0.83
≥4	0.40	0.90
≥5	0.28	0.94
≥6	0.16	0.96
≥7	0.06	0.98
≥8	0.02	0.99
≥9	0.00	1.00
= 10	0.00	1.00

children who have a PAS between 3 and 6, the score cannot accurately determine the diagnosis. This is the group in which further investigation using imaging studies is needed (Figure 1).

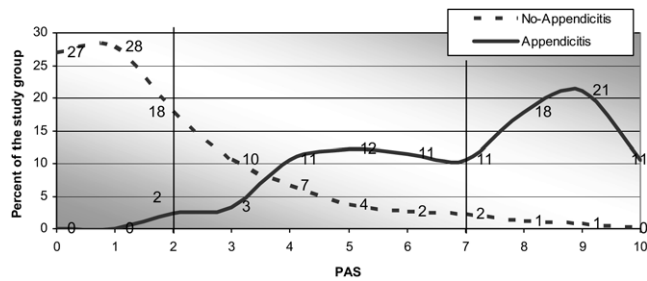
Several scoring systems have been suggested previously in the literature,<sup>12-24</sup> most in adults. In a retrospective comparison of appendectomies for appendicitis versus non-inflamed appendix,<sup>12</sup> use of a scoring system could have reduced the rate of negative appendectomies by one third. Another computerized algorithm<sup>13</sup> suggested a useful scoring system but was not tested clinically. Alvarado<sup>14</sup> found 8 predictive factors of acute appendicitis among 305 patients admitted to the hospital with abdominal pain suspected to be due to acute appendicitis. Right lower quadrant tenderness, leukocytosis, and migration of pain were the most significant factors. The study was retrospective, based only on admitted patients, with no follow-up of patients discharged with complaints of abdominal pain. The Alvarado score was later tested on 215 adult patients in Cardiff<sup>15</sup> with reduction in unnecessary



**Figure 1.** Receiver Operating Characteristic (ROC) Curve for Pediatric Appendicitis Score.

operations without documented increase in death or significant complications.

Three prospective studies compared the success of surgical teams in diagnosing appendicitis on the basis of clinical skills or clinical scores.<sup>16-18</sup> One study applied a modified Alvarado score on 84 adult females who selectively underwent laparoscopy<sup>16</sup> and resulted in zero negative explorations. In contrast, when one Australian surgical team used an 8-item score, they failed to show better diagnostic accuracy or reduction in the rate of unnecessary operations or complications.<sup>17</sup> When different surgical teams were compared in India,<sup>18</sup> a



**Figure 2.** Percent of study and control groups as a function of the Pediatric Appendicitis Score.

surgical group using a 5-criteria score had a significantly lower rate of unnecessary appendectomies ( $P < .05$ ) compared with a group not using the score. However, the sample size was only 58 patients.

Two other scores were developed on the basis of retrospective data and tested clinically and prospectively.<sup>19,20</sup> Both had low specificity and sensitivity. A 2-phase multicenter trial from Europe with more than 1400 adult patients evaluating a score prospectively<sup>21</sup> found no significant differences in the rate of perforated appendicitis, appendectomy with normal findings or complications. A reduction in delayed appendectomy and delayed discharge was documented.

Few groups challenged the creation of clinical decision scores in the pediatric population with suspected appendicitis.<sup>22-24</sup> The Alvarado score was tested in 2 studies.<sup>22,23</sup> A retrospective analysis of 156 children with a mean age of 8.4 years found that with use of the score, 9% (9/98) of patients with a diagnosis of complicated acute appendicitis were overlooked.<sup>22</sup> When examined prospectively on 187 children (2 to 17 years of age) from California,<sup>23</sup> the score failed to predict appendicitis in a significant number of patients.

In a prospective observational study from Boston, 440 patients were stratified to low, medium, and high risk of appendicitis on the basis of a recursive partitioning model.<sup>25</sup> However, the specificity for the low-risk group was only 50% and the sensitivity for the high-risk group was only 42%. Also, the group included only patients who needed a surgical consultation in the emergency department and did not follow-up on children with abdominal pain who were discharged from the emergency department without such consultation and might have had appendicitis.

Recently, Samuel<sup>24</sup> published a simple pediatric appendicitis score (PAS). During a 5-year period more than 1100 children between the ages of 4 and 15 years with abdominal pain suggestive of acute appendicitis were evaluated retrospectively in 2 hospitals in London. Eight variables were found to be significant when the group of children with appendicitis ( $n = 734$ ) was compared with the group of children without appendicitis ( $n = 436$ ). The variables were as follow: cough/percussion/hopping tenderness in the right lower quadrant (Index of 0.96), anorexia (0.88), fever (0.87), nausea or vomiting (0.86), tenderness over the right iliac fossa (0.84), leukocytosis (0.81), polymorphonuclear neutrophilia (0.80), and migration of pain (0.80). When validated on the

same patient cohort, the pediatric appendicitis score had a sensitivity of 100%, specificity of 92%, positive predictive value 96%, and negative predictive value of 99%. A score of  $\geq 6$  gave a high probability of appendicitis. However, the score has not previously been validated prospectively in a separate group of pediatric patients.

Although the PAS was developed on children 4 to 15 years old, we decided to validate it with all children over 1 year of age (1-17 years old). Although it is difficult to accurately answer questions such as whether abdominal pain has migrated and if nausea is present in children who are unable to communicate verbally, we suggested that to use the score on a large scale in primary and tertiary emergency departments, we should attempt to validate its usefulness for all age groups beyond infancy.

There are several ways in which this score may be useful. The first is in community hospitals where ultrasound and CT operators do not have substantial pediatric exposure and where a pediatric surgeon is unavailable. The score, on the basis of clinical and simple laboratory measures, is taught easily and can be used by trainees and experienced staff alike. Calculation is simple, and no special, time-consuming or invasive tests are needed. The PAS can be used as a clear communication tool between health care providers and as a guideline for transferring children from one facility to another when a pediatric surgery consultation is needed.

The second proposed use for the score is to select appropriate patients for imaging. We suggest that a child with a PAS of  $\geq 7$ , as suggested by Samuel, is very likely to have appendicitis and should be taken to the operating room without further imaging. The negative appendectomy rate of 1.8% that would result from this approach is lower than most series in the literature. A child with a PAS of 0, 1, or 2 could be observed in the emergency department or sent home with instructions to return if symptoms do not improve. Children with a PAS of 3 to 6 would undergo ultrasonography or CT. Both of these modalities have been shown to be useful in the diagnosis of appendicitis, particularly for children in whom the clinical picture is unclear. Use of the score in this way would decrease preoperative waiting time, with a potential decrease in perforation rate, and would reduce the pain associated with ultrasound scanning performed on a child with severe abdominal pain, the radiation associated with CT, and the cost associated with both studies.

Three children (2.4%) with appendicitis would have been discharged by use of the suggested cutoff point of PAS at 7. The potential saved time in the emergency department, reduced number of ultrasound examinations, and reduced exposure to CT should be further compared with the potential morbidity associated with discharge of these patients. Continuation of symptoms also might have brought these discharged patients back to the hospital.

One limitation of this study is that it was conducted in a tertiary pediatric emergency department with experienced pediatric emergency medicine (PEM) physicians and fellows. It is possible that physicians who lack experience in pediatric

emergency medicine would fail to document a similar PAS score. However, the score is simple, and only basic physical examination skills are required to allow accurate calculation of the 6 clinical components. Two other components rely on objective laboratory results. Some of the children in this study did not have blood tests drawn, and we suggest that the emergency physicians suspected these children had other reasons for abdominal pain, and not appendicitis. Also, multiple faculty and fellows collected data on our cohort, and interobserver variability should be taken into account. However, we asked only PEM fellows and faculty to collect the data, and all were trained in the data collection method.

Another limitation is associated with patients who were discharged from the emergency department but returned because of ongoing symptoms and were found to have appendicitis. Theoretically, these patients might not have been recruited the second time in the emergency department because we were unable to recruit all eligible patients. However, this is very unlikely because we were able to reach by phone almost all discharged patients for a follow-up and could verify their final diagnosis. Our series included 2 children who had appendicitis diagnosed after leaving the emergency department. They were enrolled in the "appendicitis" group. Finally, in the population recruited to the study, we had only 2 patients who underwent appendectomy, but the appendix was not found to be inflamed. This is a lower-than-expected rate of "negative appendectomy" and may represent a biased population. We did not capture all the patients arriving at the emergency department, and it is possible that some patients were evaluated, sent to the operating room, and had no appendicitis. In addition, we did not collect information on other than appendicitis pathologic findings, which could bias the results.

## REFERENCES

1. Caty MG, Azizkhan RG. Acute surgical conditions of the abdomen. *Pediatr Ann* 1994;23:192-4.
2. Aldiss DG, Shaffer N, Fowler BS, Tauxe RV. The epidemiology of appendicitis and appendectomy in the United States. *Am J Epidemiol* 1990;132:910-25.
3. Curran TJ, Muenchow SK. The treatment of complicated appendicitis in children using peritoneal drainage: results from a public hospital. *J Pediatr Surg* 1993;28:204-8.
4. McAnena OJ, Austin O, O'Connell PR, Hederman WP, Gorey TF, Fitzpatrick J. Laparoscopic versus open appendectomy: a prospective evaluation. *Br J Surg* 1992;179:818-20.
5. McLean AD, Stonebridge PA, Bradbury AW, Rainey JB, Macleod DA. Time of presentation, time of operation, and unnecessary appendectomy. *BMJ* 1993;306:307.
6. Surana R, O'Donnell B, Puri P. Appendicitis diagnosed following active observation does not increase morbidity in children. *Pediatr Surg Int* 1995;10:76-8.
7. Ohmann C, Yang Q, Franke C. Diagnostic scores for acute appendicitis. *Abdominal Pain Study Group. Eur J Surg* 1995;161:273-81.
8. Hoffmann J, Rasmussen OO. Aids in the diagnosis of acute appendicitis. *Br J Surg* 1989;76:774-9.
9. Pearl RH, Hale DA, Molloy M, Schutt DC, Jaques DP. Pediatric appendectomy. *J Pediatr Surg* 1995;30:173-8.
10. Douglas CD, Macpherson NE, Davidson PM, Gani JS. Randomised controlled trial of ultrasonography in diagnosis of acute appendicitis, incorporating the Alvarado score. *BMJ* 2000;321:919-22.
11. Roebuck DJ, Metreweli C. Radiation risk in CT for acute abdominal pain. *Radiology* 1998;209:287-8.
12. Teicher I, Landa B, Cohen M, Kabnick LS, Wise L. Scoring system to aid in diagnoses of appendicitis. *Ann Surg* 1983;198:753-9.
13. Arnbjornsson E. Scoring system for computer-aided diagnosis of acute appendicitis. The value of prospective versus retrospective studies. *Ann Chir Gynaecol* 1985;74:159-66.
14. Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med* 1986;15:557-64.
15. Owen TD, Williams H, Stiff G, Jenkinson LR, Rees BI. Evaluation of the Alvarado score in acute appendicitis. *J R Soc Med* 1992;85:87-8.
16. Lamparelli MJ, Hoque HM, Pogson CJ, Ball AB. A prospective evaluation of the combined use of the modified Alvarado score with selective laparoscopy in adult females in the management of suspected appendicitis. *Ann R Coll Surg Engl* 2000;82:192-5.
17. Gough IR. A study of diagnostic accuracy in suspected acute appendicitis. *Austr NZ J Surg* 1988;58:555-9.
18. Christian F, Christian GP. A simple scoring system to reduce the negative appendectomy rate. *Ann R Coll Surg Engl* 1992;74:281-5.
19. Izbicki JR, Knoefel WT, Wilker DK, Mandelkow HK, Muller K, Siebeck M, et al. Accurate diagnosis of acute appendicitis: a retrospective and prospective analysis of 686 patients. *Eur J Surg* 1992;158:227-31.
20. Ramirez JM, Deus J. Practical score to aid decision making in doubtful cases of appendicitis. *Br J Surg* 1994;81:680-3.
21. Ohmann C, Franke C, Yang Q. Clinical benefit of a diagnostic score for appendicitis: results of a prospective interventional study. *German Study Group of Acute Abdominal Pain. Arch Surg* 1999;134:993-6.
22. Impellizzeri P, Centonze A, Antonuccio P, Turiaco N, Cifala S, Basile M, et al. Utility of a scoring system in the diagnosis of acute appendicitis in pediatric age. A retrospective study. *Minerva Chirurgica* 2002;57:341-6.
23. 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.
24. Samuel M. Pediatric appendicitis score. *J Pediatr Surg* 2002;37:877-81.
25. Kharbanda A, Taylor G, Fishman S, Bachur R. Risk stratification in the diagnosis of appendicitis *Ped Emerg Care* 2004;20:795 [Abstract].