World J Surg (2012) 36:216–221
DOI 10.1007/s00268-011-1310-5

World Journal of Surgery

Diagnosing Appendicitis at Different Time Points in Children with Right Lower Quadrant Pain: Comparison Between Pediatric Appendicitis Score and the Alvarado Score

Han-Ping Wu · Wen-Chieh Yang · Kang-Hsi Wu · Chan-Yu Chen · Yun-Ching Fu

Published online: 19 October 2011

Abstract
Background Acute appendicitis is the most common abdominal emergency in clinical surgery. This study was designed to compare the diagnostic value of the Pediatric Appendicitis Score (PAS) with that of the Alvarado score based on different time points in children with right lower quadrant (RLQ) abdominal pain.
Methods This prospective cohort study comprised 1,395 children (aged 3–18 years) with RLQ abdominal pain between 2005 and 2009. The patients were scored by the pediatric emergency physicians. Receiver operating characteristic (ROC) curves were used to determine the appropriate cutoff scores of the Alvarado and PAS systems. The sensitivity and specificity, and area under ROC curve were calculated for both systems.
Results The AUCs of the Alvarado system were all higher than those of the PAS system (day 1: 0.09 vs. 0.87; day 2: 0.87 vs. 0.84; day 3: 0.88 vs. 0.82). The best appropriate cutoff scores were seven (days 1 and 2) and six (day 3) on the PAS and six (days 1 and 2) and seven (day 3) on the Alvarado system.
Conclusions The preliminary data show that the best cutoff score of Alvarado and PAS systems vary with the different time points of RLQ pain presentation. It may provide helpful information for primary or emergency physicians to determine whether the patient should undergo surgical consultation.

Keywords Abdominal pain · Appendicitis · Children · Gastrointestinal

H.-P. Wu · Y.-C. Fu
Institute of Clinical Medicine, National Yang-Ming University, Taipei, Taiwan, ROC

H.-P. Wu
Department of Pediatrics, Buddhist Tzu-Chi General Hospital, Taichung Branch, Taichung, Taiwan, ROC

H.-P. Wu
Department of Medicine, Tzu Chi University, Hualien, Taiwan, ROC

W.-C. Yang · C.-Y. Chen
Department of Pediatrics, Changhua Christian Hospital, Changhua, Taiwan, ROC

K.-H. Wu
Department of Pediatrics, Children's Hospital and School of Chinese Medicine, China Medical University Hospitals, Taichung, Taiwan, ROC

Y.-C. Fu (✉)
Department of Pediatrics, Taichung Veterans General Hospitals, 160, Sec. 3, Chung-Kang Road, Taichung 40705, Taiwan, ROC
e-mail: yunchingfu@gmail.com; arthur1226@gmail.com

Introduction

Appendicitis is the most frequent surgical etiology among children presenting to emergency departments (EDs) with abdominal pain over the right lower quadrant (RLQ) [1, 2]. However, despite intensive research and discussion, rapid, accurate diagnosis of pediatric appendicitis remains an elusive challenge [2–4]. A clinical decision to operate leads to the removal of a normal appendix in 10 to 20% of the cases [5, 6]. Although diagnostic imaging tools of sonography or computed tomography have been used with increasing frequency, they have limitations, such as exposure to ionizing radiation, availability of skilled technicians at all hours, and high costs [7, 8]. Moreover, some cases are not straightforward, leading to equivocal results [9–12]. Therefore, evaluation of RLQ abdominal pain should

distinguish between children with likely appendicitis who should undergo immediate surgical consultation and children with equivocal clinical presentations who should be kept under observation for further diagnostic evaluation. Classically, rapid diagnosis of acute appendicitis is commonly based on a brief detailed history, focused physical examination, and directed laboratory findings that consist of clinical diagnostic scoring systems. The Alvarado and the pediatric appendicitis score (PAS) systems have been clinically applied for diagnosing appendicitis for many years [2, 13]. Some studies recommend surgery for all patients with a score ≥ 7 and observation for patients with scores of 5 or 6 on both scoring systems [2, 7, 14]. Nevertheless, there exist differences in the weights assigned to some variables between the two scoring systems for acute appendicitis; hence, which of these diagnostic scoring systems is more appropriate for application in children with suspected appendicitis remains to be determined. Furthermore, whether the diagnostic values of the scoring systems remain effective over the various periods, from the time of onset of abdominal pain to the time of clinical examination and scoring, remains unclear. In this study, we intend to compare the diagnostic value of the PAS to that of Alvarado systems based on different time points in children with RLQ abdominal pain.

Methods

Patient participants

This study enrolled patients aged 3–18 years who presented with RLQ abdominal pain at Changhua Christian Hospital and Taichung Tzu-Chi General Hospital in central Taiwan Between July 2005 and July 2009. The duration of symptoms comprised the period from the time that the patients first felt ill until the time of admission. We designated the durations as follows: < 24 h as day 1; 24–48 h as day 2; and 48–72 h as day 3. The ultimate diagnosis of the patients who were treated surgically was based on histological examination of the excised appendix, with transmural granulocyte invasion as the histopathological criterion for the diagnosis of appendicitis. In addition, a patient was defined as having a normal appendix when a nonsurgical patient discharged from the ED was followed up by a telephone interview 2 weeks after the index visit to confirm that the diagnosis of appendicitis could be ruled out or when an uninflamed appendix was found in a patient who had undergone surgery (a normal appendectomy). The exclusion criteria included children younger than aged 3 years, patients with symptoms and signs lasting more than 3 days, and patients who were lost to follow-up at our ED or outpatient clinic. The study was approved by the institution's Human Subjects Review Committee. Informed consent was obtained from all study participants or their guardians.

Methods

This was a prospective observational cohort study and was not intended to influence the indications and timing of the surgical approach. The variables of the Alvarado scoring system include migration of pain (1 point), anorexia (1 point), nausea/emesis (1 point), tenderness in the right iliac fossa (2 points), rebound tenderness (1 point), high body temperature > 37.3°C (1 point), leukocytosis $> 10400/mm^3$ (2 points), and neutrophilia $> 75\%$ (1 point). The variables of the PAS system are as follows: 1) on history: migration of pain (1 point), anorexia (1 point), and nausea/vomiting (1 point); 2) on physical examination: fever > 38°C (1 point), cough/percussion/hopping tenderness (2 points), and RLQ tenderness (2 points); and 3) on laboratory examinations: leukocytosis $> 10000/mm^3$ (1 point) and neutrophilia $> 75\%$ (1 point). The following data were recorded on the data sheet: age; gender; body temperature; the clinical symptoms and signs of the Alvarado and PAS systems, including migration of pain, anorexia, nausea/emesis, rebound tenderness, cough/percussion/hopping tenderness, and RLQ tenderness; the time of onset of symptoms; and the time of admission. On admission, blood samples were obtained from all the selected participants, and the total white blood cell (WBC) and total neutrophil counts were measured using an automated five-part leukocyte differential count hematology analyzer (Cell-Dyn 4000R System; USA). The study data forms were completed before the imaging examinations or surgeries and the scores were analyzed and recorded carefully in a blind manner by two independent observers (the pediatric emergency physicians).

The selected participants were divided into two groups: patients with normal appendices and those with acute appendicitis. The values are presented as the mean $\pm$ standard deviation (SD) for both the scoring systems on each day. The correlations between the mean scores of the two scoring systems and the duration of development of the patients' symptoms were analyzed statistically in children with RLQ abdominal pain. In addition, we calculated all the areas under the receiver operating characteristic (ROC) curve (AUC) of the two scoring systems in discriminating acute appendicitis. Furthermore, for each scoring system, we determined two cutoff points that can more easily be applied to clinical practice to determine when the variables can rule in and rule out appendicitis. Moreover, we compared the diagnostic accuracies between the PAS and the Alvarado systems at various cutoff values for predicting acute appendicitis. Furthermore, we carefully analyzed and determined false

positives and false negatives. The definite diseases in the non-appendicitis patients group whose scores ≥ 7 of the two scoring systems also were analyzed in this study.

Statistical analysis

The *t* test, Mann-Whitney *U* test, and ROC curves were used for statistical analysis. The differences between the groups are presented as 95% confidence intervals (CIs). Probability levels < 0.05 were considered to be significant. Interobserver agreements in analyzing the scores in the two scoring systems were calculated by kappa statistics. We also examined the test parameters, including sensitivity (Sn), specificity (Sp), and AUC at the various cutoff values. The AUC, calculated using the trapezoidal rule, was considered as a global measure of the diagnostic value of that parameter. The criterion value indicated the value corresponding to the highest accuracy (minimal false-negative and false-positive results). Statistical analyses were performed by using SPSS software (version 15.0; SPSS Inc., Chicago, IL, USA).

Results

Characteristics of the study subjects

During the study period, a total of 1,562 children with RLQ abdominal pain were enrolled in the study. Of them, 167 patients who did not use our outpatient clinic for follow-up were excluded from this study. The remaining 1,395 pediatric patients aged < 18 years were recruited for further analysis; they comprised 645 boys (46.2%) and 750 girls (53.8%) with a mean age of 11.1 ± 4.2 years. Of the 1,395 cases, 881 had histologically proven acute appendicitis, and 514 had normal appendices. The final diagnoses of the patients with normal appendices included infectious enteritis, functional gastrointestinal disorders, diverticulitis, adhesion ileus, intestinal perforation, perforation of Meckel's diverticulum, mesenteric lymphadenopathy, lymphoma, and ruptured tubo-ovarian abscess.

Main results

In our study cohort, the mean scores on the Alvarado and PAS systems on days 1, 2, and 3 after symptom onset were all significantly higher in children with appendicitis than in children without appendicitis (Table 1). To derive a clinically useful estimate from the two scoring systems to predict pediatric appendicitis, the ROC curves of the PAS and Alvarado scores on days 1, 2, and 3 were plotted (Fig. 1). The ROC curves showed that both scores were highly sensitive for the diagnosis of appendicitis because their AUCs on days 1, 2, and 3 were all considerably > 0.5. The differences in the AUCs between the two scoring systems on each day are shown in Table 2. Our results revealed that both were useful scoring systems in diagnosing acute appendicitis in children with RLQ abdominal pain on days 1, 2, and 3 ($P < 0.001$ on all days); however, the AUCs of the Alvarado score in discriminating acute appendicitis in children on days 1 (0.9), 2 (0.87), and 3 (0.88) were higher than those of the PAS (day 1, 0.87; day 2, 0.84; and day 3, 0.82). Among children with a PAS ≤ 3 on days 1 and 2 and those with a PAS ≤ 4 on day 3, none had appendicitis. Additionally, among children with an Alvarado score ≤ 4 on days 1 and 2 and those with an Alvarado score ≤ 5 on day 3, none had appendicitis. Moreover, the best optimized cutoff scores of the PAS for diagnosing "appendicitis" were determined as 7 on days 1 and 2 and 6 on day 3; the best optimized cutoff scores of the Alvarado scoring system also were determined (6 on days 1 and 2 and 7 on day 3; Table 3).The overall kappa values representative of interobserver agreements showed $P < 0.01$.

Furthermore, after analysis of the non-appendicitis patients with RLQ abdominal pain, 137 patients (26.7%) had the Alvarado score ≥ 6 and 155 cases (30.2%) had the PAS ≥ 6. In the non-appendicitis group, the definite diagnoses in patients with the scores ≥ 7 of the both two clinical scoring systems consisted of infectious enteritis, diverticulitis, mesenteric lymphadenopathy, pelvic inflammatory disease, and ruptured tubo-ovarian abscess. Among the above diseases, infectious enteritis was the most common disease mimicking acute appendicitis (the Alvarado score, 17.3%; the PAS, 18.5%). On the other hand, in the appendicitis group, patients with the Alvarado score < 6 accounted for 8.6% ($n = 75$) and patients with the PAS < 6 had 13.7% ($n = 121$). Moreover, a score < 6 was more common found on day 2 (the Alvarado score, 10.2%; the PAS, 18.6%) followed by day 1 (the Alvarado score, 7.1%; the PAS, 11.7%) and day 3 (the Alvarado score, 5.1%; the PAS, 5.1%). Based on these results, the Alvarado score had lower rate of misdiagnosis than the PAS on both day 1 and day 2, but on day 3, the Alvarado score and the PAS had the same misdiagnosis rates (both 5.1%; $n = 8$).

Discussion

Some variables differ between the PAS and Alvarado scoring systems. The PAS assigns 2 points each to cough/percussion/hopping tenderness and RLQ tenderness, whereas the Alvarado scoring system assigns 2 points each to tenderness in the right iliac fossa and WBC count $>10{,}400/mm^3$. The differences in the variables and weights between the two scoring systems indicate the

Table 1 The mean Alvarado score and PAS in our study cohort of children with and without appendicitis on days 1, 2, and, 3

Scoring systems	Normal appendices Mean ± SD	Appendicitis Mean ± SD	*P* value
Day 1			
PAS	4.49 ± 2.14	7.72 ± 1.64	<0.001
Alvarado score	4.25 ± 2.07	7.79 ± 1.45	<0.001
Day 2			
PAS	4.26 ± 2.01	7.43 ± 1.88	<0.001
Alvarado score	4.57 ± 2.17	7.68 ± 1.61	<0.001
Day 3			
PAS	4.86 ± 2.33	7.48 ± 1.25	<0.001
Alvarado score	4.27 ± 2.26	7.78 ± 1.0	<0.001

Table 2 The areas under the ROC curves for the Alvarado score and PAS on days 1, 2, and 3

Diagnostic score	Area under the ROC curve			
	Area	Standard error	95% Confidence interval	*P* value
Day 1				
PAS	0.87	0.013	0.843–0.896	<0.001
Alvarado score	0.9	0.012	0.877–0.923	<0.001
Day 2				
PAS	0.84	0.022	0.794–0.879	<0.001
Alvarado score	0.87	0.023	0.821–0.909	<0.001
Day 3				
PAS	0.82	0.047	0.731–0.915	<0.001
Alvarado score	0.88	0.044	0.79–0.963	<0.001

differences in the criteria used to identify the probability of acute appendicitis. However, the decisive factor from among the different scoring systems for identifying children as those at a low or high risk for appendicitis depends on the primary clinicians and remains substantially unclear because the scoring systems do not consider the different points of time when the symptoms and signs are presented. Patients with acute appendicitis may not present with all symptoms and signs at the time of admission, and the symptoms and signs may vary with the progress of appendicitis. Therefore, the time of presentation of the symptoms and signs is important and should be considered in the clinical scoring systems for diagnosing appendicitis. Some studies have focused on children with general complaints of abdominal pain, but some have focused on children with clinically suspected appendicitis [15, 16]. The differences between the selected patient participants may have led to the differences in the diagnostic scores on the scoring systems and can indeed affect the confidence of clinicians in universal application of the clinical scoring systems for diagnosing appendicitis in clinical practice.

Table 3 Diagnostic values of the cutoff values of the Alvarado score and PAS for appendicitis on day 1 to day 3

Duration	Cutoff score	Sensitivity	Specificity	LR^+	LR^-
Day 1					
PAS	≥7	0.821	0.824	4.668	0.217
Alvarado score	≥6	0.929	0.757	3.821	0.094
Day 2					
PAS	≥7	0.713	0.827	4.122	0.347
Alvarado score	≥6	0.898	0.683	3.829	0.15
Day 3					
PAS	≥6	0.949	0.611	2.441	0.083
Alvarado score	≥7	0.949	0.778	4.272	0.065

However, to clinically rule out appendicitis, primary clinicians should more carefully evaluate children with focal RLQ pain than those with general complaints of abdominal pain. This study has determined and focused on the data at various thresholds on the PAS and Alvarado scoring systems for diagnosing pediatric appendicitis in children presenting to the ED with RLQ abdominal pain on days 1–3.

Clinically, the diagnosis of appendicitis can be divided into three diagnostic test zones: the zone with a high Sn, the zone with a high Sp, and an indeterminate zone. Based on our results, for each scoring system, we defined two cutoff points that can more easily be applied to clinical practice to determine when the variables can rule in and rule out appendicitis. Further investigation may not be needed in such patients with RLQ abdominal pain to confirm the diagnosis of appendicitis. However, primary clinicians have to pay more attention to the patients with suspected appendicitis in the indeterminate zone in clinical practice, because the decision in such patients is not easy. In our study, we have identified the optimized thresholds of the PAS and Alvarado scores for diagnosing pediatric appendicitis on days 1 to 3 in the patients in this zone. An Alvarado score ≥ 6 on days 1–2 and an Alvarado score of 7 on day 3 yielded the best diagnostic value, whereas a PAS ≥ 7 on days 1–2 and a PAS of 6 on day 3 were appropriate cutoff points in diagnosing appendicitis. In addition, we found the Alvarado score to be more valid for diagnosing acute appendicitis than the PAS in children with RLQ abdominal pain. Therefore, clinically, once the Alvarado score was greater than 6 on the first day, greater than 6 on the second day, and greater than 7 on the third day after onset of symptoms, the probability of acute appendicitis significantly increased.

However, a patient with a normal appendix suffering from RLQ pain may have an inflammatory condition in the abdomen related to something else. The cases of bowel obstruction and intestinal perforation, considered correctly in theory as false-positive cases for the diagnosis of

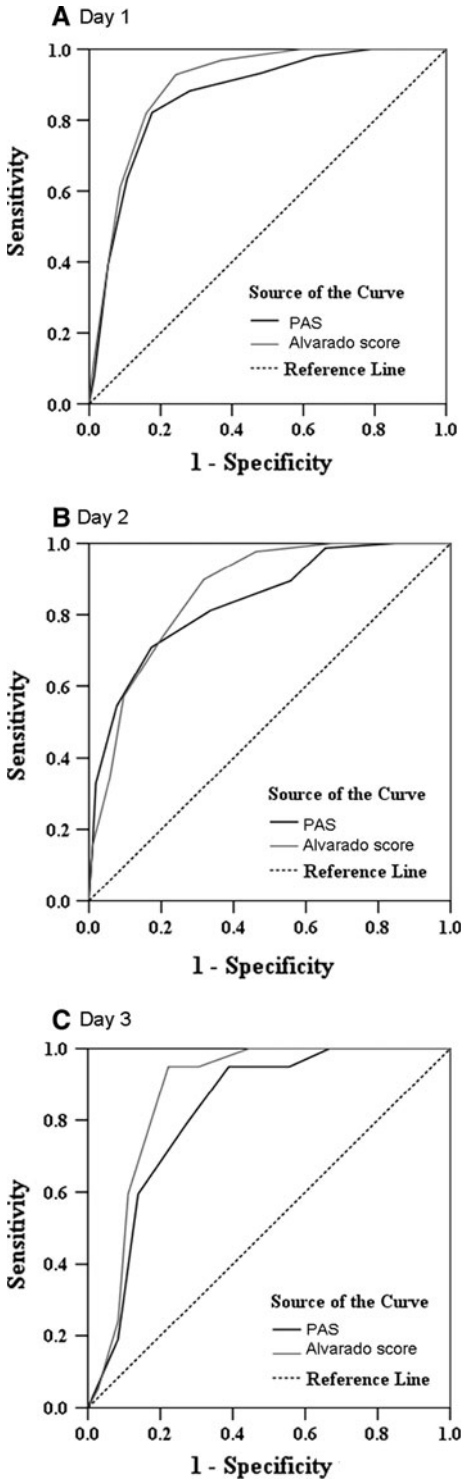

Fig. 1 ROC curves for the Alvarado score and PAS on days 1–3 (**a**–**c**, respectively)

appendicitis, cannot reflect erroneous indications for surgery, because these patients require intervention besides having a normal appendix. It also appears to be important for primary clinicians to evaluate the need of surgical consultation for children who suffer from non-appendicitis abdominal emergencies mimicking clinical presentations of acute appendicitis. Furthermore, after analyzing the non-appendicitis patients with the scores ≥ 6 of the two scoring systems, infectious enteritis was found to be the most frequent disease, which may mimic acute appendicitis in children with RLQ abdominal pain followed by functional gastrointestinal disorders. The results may highlight primary clinicians still need to pay attention for the pediatric patients with clinically suspected acute appendicitis (scores ≥ 6) to lower false-positive rate, which may easily cause negative appendectomy.

Nevertheless, there was a limitation in this study. This research is hindered by the absence of a validating set of patients subjected to appendicitis scoring using their optimized cutoff values. But in our ongoing study, we will collect another set of patients to validate the threshold scores of the scoring system established from the test set.

In conclusion, the preliminary data show that the best cutoff score of Alvarado and PAS systems vary with the different time points of RLQ pain presentation. It may provide helpful information for primary or emergency physicians to determine whether the patient should undergo surgical consultation.

Acknowledgment This study was partly supported by the grants from the China Medical University Hospital (CMU99-COL-02).

References

1. Kharbanda AB, Taylor GA, Fishman SJ, Bachur RG (2005) A clinical decision rule to identify children at low risk for appendicitis. Pediatrics 116:709–716
2. Tan EC, Severijnen RS, Rosman C et al (2006) Diagnosis and treatment of acute appendicitis in children: a survey among Dutch surgeons and comparison with evidence-based practice. World J Surg 30:512–518. doi:10.1007/s00268-005-0350-0
3. Bundy DG, Byerley JS, Liles EA et al (2007) Does this child have appendicitis? JAMA 298:438–451
4. Wu HP, Huang CY, Chang YJ et al (2006) Use of changes over time in serum inflammatory parameters in patients with equivocal appendicitis. Surgery 139:789–796
5. Kwan KY, Nager AL (2010) Diagnosing pediatric appendicitis: usefulness of laboratory markers. Am J Emerg Med 28:1009–1015
6. Bickell NA, Aufses AH Jr et al (2006) How time affects the risk of rupture in appendicitis. J Am Coll Surg 202:401–406
7. Douglas CD, Macpherson NE, Davidso PM, Gani JS (2000) Randomised controlled trial of ultrasonography in diagnosis of acute appendicitis, incorporating the Alvarado score. BMJ 321:919–922
8. Garcia Pena BM, Mandl KD, Kraus SJ et al (1999) Ultrasonography and limited computed tomography in the diagnosis and management of appendicitis in children. JAMA 282:1041–1046
9. Tsao K, St Peter SD, Valusek PA et al (2008) Management of pediatric acute appendicitis in the computed tomographic era. J Surg Res 147:221–224
10. Taylor GA, Fishman SJ, Mandl KD (2000) Costs and effectiveness of ultrasonography and limited computed tomography for diagnosing appendicitis in children. Pediatrics 106:672–676

11. Daly CP, Cohan RH, Francis IR et al (2005) Incidence of acute appendicitis in patients with equivocal CT findings. Am J Roentgenol 184:1813–1820
12. Stengel JW, Webb EM et al (2010) Acute appendicitis: clinical outcome in patients with an initial false-positive CT diagnosis. Radiology 256:119–126
13. Samuel M (2002) Pediatric appendicitis score. J Pediatr Surg 37:877–881
14. Alvarado A (1986) A practical score for the early diagnosis of acute appendicitis. Ann Emerg Med 15:557–564
15. Goldman RD, Carter S, Stephens D et al (2008) Prospective validation of the pediatric appendicitis score. J Pediatr 153:278–282
16. Schneider C, Kharbanda A, Bachur R (2007) Evaluating appendicitis scoring systems using a prospective pediatric cohort. Ann Emerg Med 49:778–784