

Predicting Acute Appendicitis? A comparison of the Alvarado Score, the Appendicitis Inflammatory Response Score and Clinical Assessment

D. Kollár · D. P. McCartan · M. Bourke ·
K. S. Cross · J. Dowdall

Published online: 23 September 2014
© Société Internationale de Chirurgie 2014

Abstract

Background Patients presenting with suspected appendicitis pose a diagnostic challenge. The appendicitis inflammatory response (AIR) score has outperformed the Alvarado score in two retrospective studies. The aim of this study was to evaluate the AIR Score and compare its performance in predicting risk of appendicitis to both the Alvarado score and the clinical impression of a senior surgeon.

Methods All parameters included in the AIR and Alvarado scores as well as the initial clinical impression of a senior surgeon were prospectively recorded on patients referred to the surgical on call team with acute right iliac fossa pain over a 6-month period. Predictions were correlated with the final diagnosis of appendicitis.

Results Appendicitis was the final diagnosis in 67 of 182 patients (37 %). The three methods of assessment stratified similar proportions (~40 %) of patients to a low probability of appendicitis ($p = 0.233$) with a false negative rate of <8 % that did not differ between the AIR score, Alvarado score or clinical assessment. The AIR score assigned a smaller proportion of patients to the high probability zone than the Alvarado score (14 vs. 45 %) but it did so with a substantially higher specificity (97 %) and positive predictive value (88 %) than the Alvarado score (76 and 65 %, respectively).

Conclusions The AIR score is accurate at excluding appendicitis in those deemed low risk and more accurate at predicting appendicitis than the Alvarado score in those deemed high risk. Its use as the basis for selective CT imaging in those deemed medium risk should be considered.

Introduction

The estimated lifetime incidence of acute appendicitis is approximately 7 % [1]. Despite this, the majority of patients who present with acute right iliac fossa (RIF) pain do not have appendicitis and the scenario continues to pose a diagnostic challenge [2]. The Alvarado Score is the best known clinical prediction rule for estimating risk of

appendicitis [3]. It is based on a combination of symptoms, signs and basic laboratory results and has been the subject of many validation studies [4, 5]. Its use in routine clinical practice is varied and limitations including overestimating the risk of appendicitis in women and children have been noted [6].

The Appendicitis Inflammatory Response score (AIR) is based along the same principles of the Alvarado score, assigning patients to low, medium or high probability of acute appendicitis [7]. It incorporates CRP as a variable in the score, a widely available laboratory test that has not shown sufficient sensitivity or specificity to be used as a stand-alone test to predict risk of appendicitis [8]. Initially

D. Kollár · D. P. McCartan (✉) · M. Bourke ·
K. S. Cross · J. Dowdall
Department of Surgery, Waterford Regional Hospital, Dunmore
Road, Waterford, Republic of Ireland
e-mail: damiamccartan@rcsi.ie

developed on a retrospective cohort, the AIR score has been the subject of one retrospective validation study that confirmed an improved discriminating power when compared to the Alvarado score [9].

The aim of this study was to evaluate the AIR score and compare its performance in predicting the risk of appendicitis to both the Alvarado score and the clinical impression of a senior surgeon.

Materials and methods

All patients referred to the General Surgical team on call with acute RIF pain from 1st January to 1st July 2013 were

Table 1 Comparison of parameters included in Alvarado and appendicitis inflammatory response (AIR) Scores

	Alvarado score	AIR score
Symptoms		
Nausea/vomiting	1	–
Vomiting	–	1
Anorexia	1	–
RIF pain	2	1
Migratory pain to RIF	1	–
Signs		
Rebound tenderness	1	–
Muscular defence/rebound light	–	1
Medium	–	2
Strong	–	3
Temperature		
>37.3 °C	1	–
>38.5 °C	–	1
Lab results		
WBC count		
> 10.0 × 10 ⁹ /l	2	
10.0–14.9 × 10 ⁹ /l		1
≥ 15.0 × 10 ⁹ /l		2
PMN leucocytes		
> 75 %	1	–
70–84 %	–	1
≥ 85 %	–	2
CRP concentration		
10–49 g/l	–	1
≥ 50 g/l	–	2
Total score		
Low risk/probability appendicitis	0–4	0–4
Medium risk/probability appendicitis	5–6	5–8
High risk/probability appendicitis	7–10	9–12

Original references: Alvarado Score [3], AIR score [7]

included. Data were collected prospectively on a proforma completed by the surgical resident who assessed the patient in the Emergency Department. All parameters included in both the Alvarado and AIR scores were documented (Table 1).

A consultant surgeon or senior resident (post graduate year 5 or greater) was asked to categorise each patient into either a low, medium or high risk probability group for appendicitis based on their initial assessment. The findings of any operative interventions were recorded, and a diagnosis of appendicitis was based on the histological finding of transmural inflammation. Advanced appendicitis was defined as the presence of either transmural gangrene or perforation on final histology.

At present, no clinical scoring system for estimating risk of appendicitis is used within the department. Decisions regarding radiological investigations or surgery are at the discretion of the consultant surgeon in charge of the care of the patient. The Alvarado and AIR scores were calculated and patients stratified into the relevant high, medium and low risk groups at the end of the study period. The original Alvarado score separates those with a likely diagnosis of appendicitis into ‘probably appendicitis (score 7 or 8)’ and ‘highly likely appendicitis (score 9 or 10)’ yet recommended surgery for all with a score of over 7 [3]. For the purposes of this analysis, any patient with an Alvarado score ≥7 was considered as ‘High probability of appendicitis’ to allow comparison with the AIR score and the clinical judgement both of which stratified to three groups.

Statistical analysis was performed using Minitab statistical software (Minitab Inc, PA, USA). Categorical variables were assessed using Fishers exact test. A *p* value of less than 0.05 was considered significant. The diagnostic performances were assessed on calculated values for sensitivity, specificity, positive and negative predictive values as well as area under the receiver operating characteristic (ROC) curves. Differences in sensitivity and specificity were calculated using the McNemar test.

Results

A total of 182 patients with a median age of 19 years were included (Table 2). Appendicitis was the final diagnosis in 37 % of patients. Only 10 patients over the age of 16 did not undergo either radiological imaging or surgery. Of patients proceeding to surgery, 86 % underwent a laparoscopic procedure.

The Alvarado score assigned a higher proportion of patients to the high probability for appendicitis group (45 %) than the AIR score (14 %) or clinical assessment by a senior surgeon (29 %) (*p* < 0.001) (Table 3). In those deemed high probability of appendicitis, the AIR score had

Table 2 Basic demographics and outcomes of study cohort

Characteristic	Total 182 (%)
Gender	
Male	79 (43 %)
Female	103 (57 %)
Age (Median) [Range]	19.3 years [4–75]
<16 years	65 (36 %)
>16 years	117 (64 %)
Final diagnosis appendicitis	67 (37 %)
Negative appendectomy	12 (16 %)
Imaging	
Ultrasound	59 (32 %)
CT	30 (17 %)
No imaging	93 (51 %)
Operative intervention	88 (48 %)
Open appendectomy	9 (10 %)
Laparoscopic appendectomy	62 (71 %)
Laparoscopic converted to open appendectomy	3 (3 %)
Diagnostic laparoscopy	14 (16 %)

considerably higher specificity (97 %) and positive predictive values (88 %) than the Alvarado score for predicting a diagnosis of appendicitis (Table 4).

Assessment by either scoring system or clinical judgement alone assigned similar proportions of patients to the low probability of appendicitis group ($p = 0.615$). The number of patients in the low risk group who had a final diagnosis of appendicitis was 8 % or less with the scoring systems and with clinical assessment. While the AIR score had the lowest false negative rate, there was no statistical difference ($p = 0.784$) (Table 3). The percentage of patients who did not have appendicitis and were correctly assigned to the low risk group was 67 % for surgeon assessment, 62 % for the AIR score and 55 % for the Alvarado score ($p = 0.164$).

No patient with advanced appendicitis was stratified as low probability by either the AIR or Alvarado score. While not of statistical significance ($p = 0.502$), two patients deemed low probability of acute appendicitis by clinical assessment had a final diagnosis of advanced appendicitis proving that use of a scoring system is at least equally as good as clinical assessment at identifying those patients with advanced appendicitis (Table 4). All three methods of assessment stratified a similar, high proportion (≥ 75 %) of those with advanced appendicitis to the high probability of appendicitis group ($p = 0.350$). Of those deemed high probability of appendicitis by the AIR score, 48 % had a final diagnosis of advanced appendicitis, a figure higher than that for the Alvarado (19 %) and clinical assessment by a senior surgeon (25 %) ($p = 0.012$).

Radiological imaging was used as a diagnostic adjunct in 49 % of patients with ultrasound employed twice as frequently as CT. The decision to order imaging was based on clinician assessment and not the AIR or Alvarado score. Those deemed high probability for appendicitis were more likely to undergo a CT ($p = 0.003$) than those deemed less likely to have appendicitis. Overall, ultrasound had a false negative rate of 14 % for acute appendicitis that did not differ according to surgeon stratified risk of appendicitis.

Comparison of area under the curve (AUC) values from analysis of receiver operator characteristic (ROC) curves (Table 5) shows that the discriminatory capacities of all the three methods of assessment are similar with almost identical ROC curves for the AIR and Alvarado scores when plotted across all data points in each scoring system (AUC: 0.850 AIR Score, 0.84 Alvarado score). The AIR score performed better in children and in men. Assessment by a senior surgeon performed as well as both scoring systems in children and adults and had greater discriminatory capacity when assessing men but showed greatest inaccuracy when assessing females. The Alvarado score had the greatest discriminatory capacity in females.

Table 3 Patients grouping and correlation with final diagnosis of appendicitis as stratified by AIR score, Alvarado score and senior surgeon assessment

Categorised as	Scoring system	Patients Categorised n (%)	Final diagnosis Appendicitis n (%)	Advanced Appendicitis n (%)
High risk	AIR score	25 (14)	22 (88)	12 (48)
	Alvarado score	81 (45)	53 (65)	15 (19)
	Senior surgeon	52 (29)	42 (81)	13 (25)
Medium risk	AIR score	82 (45)	41 (50)	4 (5)
	Alvarado score	33 (18)	9 (27)	1 (3)
	Senior surgeon	46 (25)	18 (40)	1 (2)
Low risk	AIR score	75 (41)	4 (5)	0 (0)
	Alvarado score	68 (37)	5 (7)	0 (0)
	Senior surgeon	84 (46)	7 (8)	2 (2)

Table 4 Comparison of sensitivity, specificity, positive and negative predictive values of AIR score, Alvarado score and senior surgeon assessment for predicting acute appendicitis

	High probability appendicitis			Low probability appendicitis		
	AIR	Alvarado	Senior surgeon	AIR	Alvarado	Senior surgeon
All appendicitis						
Sensitivity	0.33	0.79 ^a	0.63 ^b	0.94	0.93	0.90
Specificity	0.97	0.76 ^a	0.91	0.62	0.55	0.67
PV+	0.88	0.65	0.81	0.59	0.54	0.61
PV–	0.71	0.86	0.81	0.95	0.93	0.92
Advanced appendicitis						
Sensitivity	0.75	0.94	0.81	1.00	1.00	0.88
Specificity	0.92	0.60 ^a	0.77 ^b	0.45	0.41	0.49
PV+	0.48	0.19	0.25	0.15	0.14	0.14
PV–	0.97	0.99	0.98	1.00	1.00	0.97

Results of the McNemar comparing the sensitivity and specificity of the AIR score, Alvarado score and clinical assessment

^a Denotes $p < 0.05$ when AIR score compared to Alvarado score

^b Denotes $p < 0.05$ when AIR score compared to clinical assessment

All other results not marked represent a result of $p > 0.05$ on McNemar test

Table 5 Discriminatory capacity of the AIR score, Alvarado score and senior surgeon assessment according to patient age and gender using receiver operator characteristic (ROC) curve analysis

	No. of patients	AIR score	Alvarado score	Senior surgeon
Overall	182	0.824	0.863	0.863
Age				
<16 years	65	0.873	0.883	0.892
>16 years	117	0.805	0.853	0.844
Sex				
Male	79	0.853	0.804	0.932
Female	103	0.795	0.883	0.756

Values presented are for area under the curve (AUC)

Discussion

This is the first evaluation of the AIR score to compare its ability to estimate risk of appendicitis with the Alvarado score and the clinical impression of the senior assessing surgeon. The results show that both the AIR and Alvarado scores are accurate in ruling out appendicitis in those stratified as low risk with high negative predictive values. The ability of the scoring systems to accurately rule out appendicitis was equal to the clinical judgement of a senior surgeon suggesting that the scoring systems are well placed to be used as a decision support tool for junior surgeons or emergency department doctors when evaluating patients with a low probability of appendicitis who could be safely selected for observation on an outpatient basis.

Important differences do exist between the AIR and Alvarado score when it comes to selecting those at high probability for acute appendicitis. A high AIR score has excellent specificity and positive predictive values that exceed those of the Alvarado score and the clinical impression of a senior surgeon. A 2011 systematic review of the Alvarado score recorded a pooled specificity of 82 % for those with a high Alvarado score for accurate prediction of appendicitis, similar to the 76 % reported here [6]. This low specificity is one factor that may explain the low rates of utilisation of the Alvarado score in clinical practice. These prospective results evaluating the specificity of a high AIR score mirror those of the original study [7] and the validation study published in 2012 [9] that demonstrate a markedly superior specificity for a high AIR score in predicting acute appendicitis. A high AIR score does not identify all patients with appendicitis but in those it does deem high risk it does so with substantially greater accuracy than the Alvarado score with a specificity upwards of 95 %, a level that should allow confidence in the test in a clinical setting.

Many aspects of acute appendicitis have changed in the last 20 years with the advent of minimal access surgery as a diagnostic and a therapeutic procedure as well as increasing use of computed tomography (CT) in achieving a pre-operative diagnosis [10]. CT has a sensitivity rate of over 95 % and its use has helped reduce negative appendectomy rates [11–13]. However, the liberal use of CT has raised concerns regarding unnecessary exposure to ionising radiation [14]. It is unrealistic to expect a clinical scoring system to achieve the same degree of sensitivity and

specificity associated with CT. Symptoms and signs associated with appendicitis frequently overlap with other inflammatory conditions such as terminal ileitis, mesenteric adenitis and pelvic conditions that can affect women of child-bearing age.

The low sensitivity of a high AIR score (33 %) is in keeping with that from the two retrospective studies (37 %) and reflects the high proportion of patients with appendicitis stratified to the medium probability group [7, 9]. This need not be construed as a failure of the AIR score and could help rationalise the use of imaging. The AIR score more confidently identifies those patients with a high probability of appendicitis in whom supplemental imaging is unlikely to change management. Selective radiological imaging may be better reserved for patients deemed medium probability from the AIR score.

It is not clear why clinical prediction rules for acute appendicitis have not attained routine clinical use, especially as other areas of medicine have assimilated clinical scoring systems into daily practice. The use of the wells score in estimating risk of deep venous thrombosis [15] and the CHADS2 scoring system [16] in predicting risk of stroke in patients with atrial fibrillation have both been developed within the last 20 years yet are commonly used. The prevailing use of smartphone and app technology should mean that these scoring systems do not have to be committed to memory but should be easily accessible. The AIR score could be incorporated as a routine diagnostic tool for use by junior clinicians in the disciplines of Emergency Medicine and General Surgery who are frequently faced with patients with suspected appendicitis who have yet to appreciate some of the nuances learned by senior colleagues through years of experience in assessing such patients. The findings of this study show that a low AIR score is as accurate in excluding appendicitis, and a high score more accurate at predicting appendicitis, than the clinical assessment of a senior surgeon.

Admittedly, this study was confined to one surgical department (seven consultants and seven senior residents) and the figures for the diagnostic accuracy of clinical judgement cannot be extrapolated to other departments. However, multiple studies have shown that while clinical assessment is not 100 % accurate, it does have a role in estimating risk of appendicitis and that a slide towards routine use of CT for patients with suspected appendicitis is not only unnecessary but also potentially harmful [17, 18].

In conclusion, the AIR score is a reproducible assessment tool that is accurate in excluding appendicitis in those deemed low probability for appendicitis. A high specificity, greater than that of the Alvarado score, make it well placed

to be used as a decision support tool in identifying patients at high probability of appendicitis that should proceed to surgery. A randomised control trial should be considered to study the AIR score as grounds for selective use of CT in those deemed medium probability for appendicitis.

Acknowledgments We wish to acknowledge the invaluable assistance of Dr Michael Harrison (Waterford Institute of Technology) for the statistical support for calculation of ROC curves and McNemars test.

Conflicts of interest The authors have no conflicts of interest to declare.

References

1. Addiss DG, Shaffer N, Fowler B et al (1990) The epidemiology of appendicitis and appendectomy in the United States. *Am J Epidemiol* 132:910–925
2. McCartan DP, Fleming FJ, Grace PA (2010) The management of right iliac fossa pain—is timing everything? *Surgeon* 8:211–217
3. Alvarado A (1986) A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med* 15:557–564
4. Owen TD, Williams H, Stiff G et al (1992) Evaluation of the Alvarado score in acute appendicitis. *J R Soc Med* 85:87–88
5. Ohmann C, Yang Q, Franke C (1995) Diagnostic scores for acute appendicitis. Abdominal Pain Study Group. *Eur J Surg* 161:273–281
6. Ohle R, O'Reilly F, O'Brien KK et al (2011) The Alvarado score for predicting acute appendicitis: a systematic review. *BMC Med* 9:139
7. Andersson M, Andersson RE (2008) The appendicitis inflammatory response score: a tool for the diagnosis of acute appendicitis that outperforms the Alvarado score. *World J Surg* 32:1843–1849. doi:10.1007/s00268-008-9649-y
8. Yu C-W, Juan L-I, Wu M-H et al (2013) Systematic review and meta-analysis of the diagnostic accuracy of procalcitonin, C-reactive protein and white blood cell count for suspected acute appendicitis. *Br J Surg* 100:322–329
9. De Castro SMM, Ünlü C, Steller E et al (2012) Evaluation of the appendicitis inflammatory response score for patients with acute appendicitis. *World J Surg* 36:1540–1545. doi:10.1007/s00268-012-1521-4
10. Sporn E, Petroski GF, Mancini GJ et al (2009) Laparoscopic appendectomy—is it worth the cost? Trend analysis in the US from 2000 to 2005. *J Am Coll Surg* 208:179–185
11. Anderson BA, Salem L, Flum DR (2005) A systematic review of whether oral contrast is necessary for the computed tomography diagnosis of appendicitis in adults. *Am J Surg* 190:474–478
12. Krajewski S, Brown J, Phang PT et al (2011) Impact of computed tomography of the abdomen on clinical outcomes in patients with acute right lower quadrant pain: a meta-analysis. *Can J Surg* 54:43–53
13. Rao PM, Rhea JT, Novelline RA et al (1998) Effect of computed tomography of the appendix on treatment of patients and use of hospital resources. *N Engl J Med* 338:141–146
14. Shah DJ, Sachs RK, Wilson DJ (2012) Radiation-induced cancer: a modern view. *Br J Radiol* 85:e1166–e1173
15. Wells PS, Hirsh J, Anderson DR et al (1995) Accuracy of clinical assessment of deep-vein thrombosis. *Lancet* 345:1326–1330

16. Gage BF, van Walraven C, Pearce L et al (2004) Selecting patients with atrial fibrillation for anticoagulation: stroke risk stratification in patients taking aspirin. *Circulation* 110:2287–2292
17. Gwynn LK (2001) The diagnosis of acute appendicitis: clinical assessment versus computed tomography evaluation. *J Emerg Med* 21:119–123
18. Hong JJ, Cohn SM, Ekeh AP et al (2003) A prospective randomized study of clinical assessment versus computed tomography for the diagnosis of acute appendicitis. *Surg Infect (Larchmt)* 4:231–239