

# Comparison of the use of Alvarado and AIR scores as an adjunct to the clinical diagnosis of acute appendicitis in the pediatric population

Aya Musbahi,<sup>OR</sup> Darren Rudd, Matei Dordea, Bussa Gopinath, Vijay Kurup

**To cite:** Musbahi A, Rudd D, Dordea M, *et al.* Comparison of the use of Alvarado and AIR scores as an adjunct to the clinical diagnosis of acute appendicitis in the pediatric population. *World Jnl Ped Surgery* 2019;2:e000040. doi:10.1136/wjps-2019-000040

Received 17 January 2019

Revised 22 April 2019

Accepted 24 April 2019

## ABSTRACT

**Background** Acute appendicitis is one of the most common causes of acute abdominal pain with an incidence of 1.17 per 1000 and lifetime risk of approximately 7%. It remains the most common indication for emergency abdominal surgery in childhood. Diagnosis of acute appendicitis is particularly difficult in young women and the pediatric population. In the USA, CT imaging is used to avert diagnostic dilemma, however the procedure is associated with radiation risk in this vulnerable population. Additionally, the procedure has high cost and variable availability.

**Methods** A retrospective study involving all suspected pediatric cases of appendicitis between the ages of 5 and 17 who were operated on between 2012 and 2015 was carried out. Data were collated from clinical notes on age, sex, ultrasound findings; postoperative complications, white cell count, neutrophils, C-reactive protein, histology result, and number of days to theater. All patients in the time period were retrospectively scored on the Alvarado and Appendicitis Inflammatory Response (AIR) scores.

**Results** A total of 239 patients between 11 and 17 (mean 13.6±SE) years of age were included in the study. Of these, 79 had preoperative ultrasound, of which 52 were negative, and only one patient had CT scan. 213 of the patients had an appendectomy and 26 had diagnostic laparoscopy with no appendectomy. Of the 213 appendixes removed, 71 were histopathologically normal, giving a negative appendectomy rate of 33.3%. 28 appendixes were perforated. The average number of days from admission to theater was 1.0 SE in males and 1.424 in females ( $p=0.0498$ ). The average number of days from admission to theater in those who had ultrasound was 2.03 days compared with 0.75 in those who did not have ultrasound ( $p<0.0001$ ). AIR scoring that was high and medium risk showed slightly lower negative appendectomy rates but not significantly different.

**Conclusions** Our study has found no significant difference between the AIR scores and Alvarado. There is a role for scoring systems to be used to aid in the decision to undergo imaging and as an adjunct to clinical diagnosis.

## INTRODUCTION

Acute appendicitis is one of the most common causes of acute abdominal pain with an incidence of 1.17 per 1000 and lifetime risk

of approximately 7%.<sup>1</sup> It remains the most common indication for emergency abdominal surgery in childhood. Diagnosis of acute appendicitis is particularly difficult in young women and the pediatric population. In the USA, CT imaging is used to avert diagnostic dilemma, however the procedure is associated with radiation risk in this vulnerable population. Additionally, the procedure has a high cost and variable availability.<sup>2</sup> Ultrasound use is common as an alternative, however it may be highly operator dependent and has variable reported sensitivity and specificity in both the pediatric and adult population.<sup>3</sup>

Scoring systems have been designed to aid clinical assessment of acute appendicitis; the Alvarado score being the most well known,<sup>4</sup> However, the Alvarado score may overpredict diagnosis of acute appendicitis in the pediatric population, contributing to increased negative appendectomy rates and leading to unnecessary morbidity and even mortality.<sup>5</sup> The Appendicitis Inflammatory Response (AIR) score<sup>6</sup> has been found to outperform the Alvarado score in retrospective studies in the adult population.<sup>7</sup>

In this study we compare the Alvarado and AIR scores in a pediatric population with a clinical diagnosis of appendicitis. At our District General Hospital, no scoring system and no guideline or protocol is formally used for diagnosing appendicitis. Management is at the discretion of the senior clinician, a consultant general surgeon. There is a need to establish how a scoring system would improve provision of appropriate treatment for appendicitis.

## METHODS

A retrospective study involving all suspected pediatric cases of appendicitis between the ages of 5 and 17 who were operated on between 2012 and 2015 was carried out. Data



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

University Hospital of North Tees, Stockton-on-Tees, UK

## Correspondence to

Dr Aya Musbahi; musbahiaya@me.com

**Table 1** Alvarado and AIR scoring systems and risk grouping

Variables	Alvarado	AIR
Nausea and/or vomiting	1	–
Vomiting	–	1
Anorexia		–
RIF pain		1
Migratory pain to RIF		–
Rebound tenderness	1	–
Rebound light	–	1
Medium	–	2
Strong	–	3
Temperature >37.3	1	–
Temperature >38.5	–	1
WCC >10×10 <sup>9</sup> /L	2	–
WCC 10–14.9×10 <sup>9</sup> /L	–	1
WCC ≥15×10 <sup>9</sup> /L	–	2
PMN leucocytes		
>75%	1	–
70%–84%	–	1
≥85%	–	2
CRP		
10–49	–	1
≥50	–	2
Total score		
Low risk/probability	0–4	0–4
Medium risk/probability	5–6	5–8
High risk/probability	7–10	9–12

AIR, Appendicitis Inflammatory Response; CRP, C-reactive protein; PMN, polymorphonuclear neutrophils; RIF, Right iliac Fossa; WCC, white cell count.

were collated from clinical notes on age, sex, ultrasound findings; postoperative complications, white cell count, neutrophils, C-reactive protein (CRP), histology result, and number of days to theater. All patients in the time period were retrospectively scored on the Alvarado and AIR scores.

The Alvarado score stratification was modified slightly based on Kollár *et al.*'s method<sup>8</sup> of predicting acute appendicitis in adults to match the AIR scoring system, which provides a similar stratification to that used by Ohle *et al.*<sup>9</sup> Patients were stratified into low, medium and high risks of appendicitis groups (table 1).

Differences between the scoring systems in their assignment to different levels of risk, and the associated concordance of the assessment with surgical observations were analyzed using two approaches. First, the assignment of patients to risk categories by the Alvarado and AIR Diagnostic methods was directly compared using Fisher's exact test of the two-way table Diagnostic\*Level. Second, proportions of different binary outcomes

**Table 2** Demographic table

Demographic variable	Patients (n=239)
Age	13.56 (11–17)
Sex	
Male	120
Female	119
Ultrasound	
Yes	79
No	159
Normal ultrasound	52
Free fluid	27

(positive vs negative appendix removal) were tested with the log-linear model:

Outcome (1/0)=Diagnostic+Level+Diagnostic\*Level

The model was calculated in R glm (), with type II hypothesis tests generated from the car: Anova() library.

## RESULTS

A total of 239 patients between 5 and 17 (mean 13.6±SE) years of age were included in the study (table 2). Of these, 79 had a preoperative ultrasound, of which 52 were negative, and only one patient had a CT scan. Two hundred and thirteen of the patients had an appendectomy and 26 had diagnostic laparoscopy with no appendectomy. Of the 213 appendixes removed, 71 were histopathologically normal, giving a negative appendectomy rate of 33.3%. Twenty-eight appendixes were perforated. The average number of days from admission to surgery was 1.0±SE in males and 1.424 in females (p=0.0498). The average number of days from admission to surgery in those who had ultrasound was 2.03 days compared with 0.75 in those who did not have ultrasound (p<0.0001). Table 3 highlights operative findings and complications.

The two classification methods did not provide similar allocation to risk categories (Fisher's exact test p<0.001). This was primarily due to the very low (five individuals) allocation to the high-risk category by AIR (figure 1). Proportionally, however, there was no indication of a significant interaction between the diagnostic method and the risk level (table 4, Diagnostic\*Level p=0.378), indicating that within a risk level, the two methods did not significantly differ. Risk level was the strongest factor accounting for variability in appendicitis rates (table 4). The low-risk category had negative appendectomy rates of 63.3% and 55.8% for Alvarado and AIR, respectively. The negative rate of the mid-risk category was considerably lower (29% and 14.9% for Alvarado and AIR, respectively).

even in the high-risk category, however, there was still an 11.5% negative rate for the alvarado diagnostic, and one out of five individuals for air (figure 1).

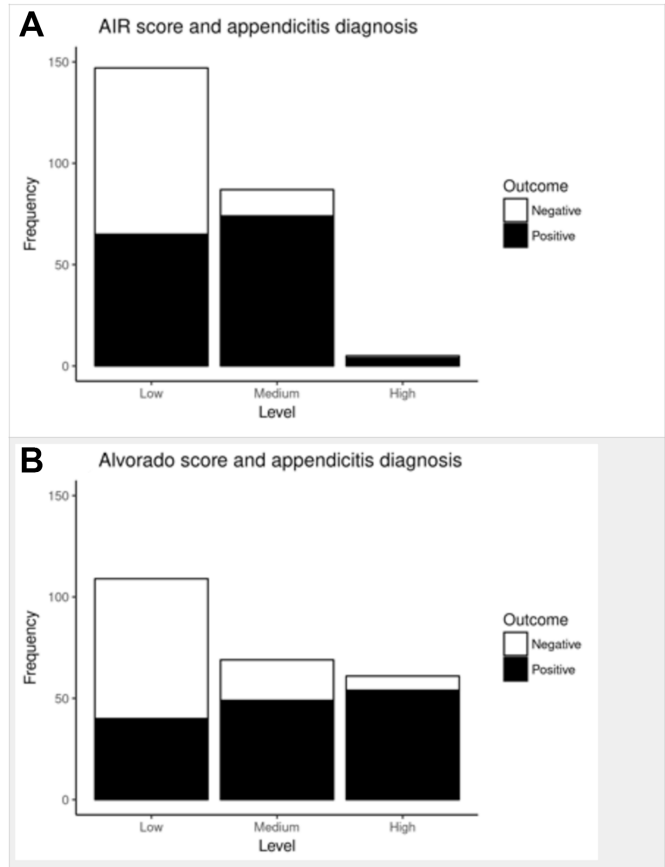
**Table 3** Operative findings and complications

Variables	n	P value
Operative findings		
Appendectomy	213	
Appendix not removed	26	
Perforated appendix	28	
Histopathological findings		
Normal	71	
Appendicitis	137	
Neuroendocrine tumor	2	
Pinworms	3	
Negative appendectomy rate	33.33%	
Average days to surgery		
Males	1.008	
Females	1.424	0.0498
Ultrasound	2.03	
No ultrasound	0.75	<0.0001
Postoperative complications		
Wound infection	7	
Pelvic collection	7	
Complications related to negative appendectomies		
Wound infection	3	
Pelvic collection	1	

## DISCUSSION

Ideally, a scoring system would work as a tool alongside clinical acumen to increase the accuracy of decision-making, while reducing the need to expose patients to harmful imaging and/or increased time before surgical intervention to prevent appendicular perforation. Previous systematic reviews have documented poor specificity of the Alvarado in pediatric populations.<sup>10-12</sup> Similarly, in comparative studies, the AIR method was superior in terms of specificity, but not necessarily sensitivity in the adult population.<sup>13</sup> This is the first study to compare the AIR and Alvarado scoring systems exclusively within a pediatric population who already had a clinical diagnosis of appendicitis and a decision to operate. The aim was to find out whether scoring systems alone can reduce the negative appendectomy rate. In contrast with other studies,<sup>14-16</sup> we did not find AIR to be superior to Alvarado in specificity in our population.

The low-risk classification classes of both AIR and Alvarado systems were similarly uninformative about appendicitis rates. However, both systems performed better at medium and high-risk levels. Adopting either or both scoring systems as a supplemental decision-making tool must take this into account with risk assessment. If in doubt, then a score of medium or high risk in either or both of the scoring systems should favor early surgical



**Figure 1** Relative frequencies of classification into low, medium and high-risk categories of (A) AIR and (B) Alvarado scoring systems. Open fills are negative appendicitis rates. AIR, Appendicitis Inflammatory Response.

intervention. There is no indication that there is clinically important distinction between medium and low-risk classes. In AIR, only five cases were assigned to the high-risk category. In the Alvarado system, although negative appendectomy rates were lower in high versus medium-risk classes, the advantage in decrease in negative rate may be outweighed by increase in danger due to perforation. A low-risk classification within either scheme will result in >50% negative appendectomy rates.

**Table 4** Analysis of variance of appendicitis occurrence as a function of classification to risk categories (Level) by different Diagnostic types (Alvarado and AIR)

Source	SS	df	F	P value
Diagnostic	4.30	1	4.248	0.0399
Level	92.02	2	45.434	<2e-16
Diagnostic*Level	1.97	2	0.974	0.378
Residual	478.00	472		

The response variable is a binary measure of positive versus negative appendectomy results. The model is a generalized linear model with binomial response errors. Tests are type II sums of squares.

AIR, Appendicitis Inflammatory Response.

Supplementary imaging is a common diagnostic tool guiding decision to operate. However, imaging as a matter of course has two main disadvantages. Ultrasound is variously advocated. Pershad *et al* found ultrasound was the most cost-effective diagnostic approach in children with suspected appendicitis,<sup>17</sup> while others have reported ultrasound imaging as inappropriate given its delay on treatment.<sup>18</sup> In this study we found that time to theater was considerably extended with ultrasound (2.03 vs 0.75 days). In reality, however, management strategies are rarely based on negative ultrasound, offsetting some of its benefit except as a confirmatory tool. However, the sonographer may also find other signs such as free fluid and thickened bowel loops, but their sensitivity and specificity in diagnosing appendicitis is uncertain.<sup>19</sup> CT imaging has been shown to have significantly higher sensitivity than ultrasound in the pediatric population,<sup>20</sup> but increases risk of radiation exposure. Our negative appendectomy rates were higher than reported in the literature. This may be due to non-reliance and reluctance on imaging such as CT within our hospital. Adoption of a scoring system could rationalize the need for selective radiological assessment in cases where the insight into equivocal cases in the low-risk category outweighs the risk of the procedure and unnecessary surgery. Radiological scans in medium and high-risk categories would be largely redundant.

The AIR score was found to more confidently identify those patients with a high probability of appendicitis in whom supplemental imaging is unlikely to change management and thus an early decision to operate should be made. This is of benefit as imaging is also shown to increase time to theater.

Our study has some limitations and caveats. Ours was a retrospective study of all children who were operated for suspected acute appendicitis. The preselected population created an inherent systematic bias. Given some degree of clinical acumen had already been used to rule out an unknown percentage of cases, this study does not attempt to generalize the performance data across a pediatric population presenting with non-specific abdominal pain, for which appendicitis may be a differential diagnosis. Our study population is clearly of those within the age range for whom appendicitis has been diagnosed.

Due to the nature of retrospective data collection, there were instances where scores had to be inferred, based on the accuracy of the documentation, or classified as 'not present/negative' if there was no documentation. Rebound tenderness when documented was graded as 1 point but no grade given when not documented. 'Moderate' rebound tenderness was given 2 points whereas descriptions such as 'severe' rebound tenderness were given 3 points. One of the benefits of the AIR score over the Alvarado score became apparent. The Alvarado score requires the child to be able to recognize and describe feelings such as nausea and verbalize migratory pain and recognize anorexia. The AIR relies more on laboratory findings and objective clinical findings.

Symptomology and signs of appendicitis are relative to the progression of the inflammatory response and time of presentation. In our data collection, we decided to look at results on admission, therefore, a selection of patients may have benefited from a scoring tool used on admission and repeatedly during observation prior to discharge.

Given the clinical overlap between acute appendicitis and non-surgical causes of abdominal pain, a scoring system that stratifies risk of acute appendicitis certainly does have a place. In high scoring children, the decision should be made to operate with no need for imaging to confirm or disprove diagnosis. There is perhaps a requirement for a modified AIR score for use in the pediatric population which could draw on observations of serum CRP and percentage polymorphonuclear leucocytes being directly related to disease stratification while remaining clear of criteria difficult for children to appreciate.

## CONCLUSION

Our study has found that diagnostic scoring systems such as the AIR and Alvarado can be used as an adjunct to clinical diagnosis to decide on whether to use imaging and reduce negative appendectomy rate, yet reduce time to theater for the majority of cases.

**Contributors** AM and DR wrote the paper. MD, VK and BG edited and supervised the paper.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## REFERENCES

1. Ergul E. Importance of family history and genetics for the prediction of acute appendicitis. *Internat J Surg* 2007;10.
2. Doria AS, 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.
3. Reddan T, Corness J, Mengersen K, *et al*. Ultrasound of paediatric appendicitis and its secondary sonographic signs: providing a more meaningful finding. *J Med Radiat Sci* 2016;63:59–66.
4. Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med* 1986;15:557–64.
5. Arif AS, Amin S, Quamruzzaman SM, *et al*. Diagnosing acute appendicitis in children using Alvarado score. *Akmmc J* 2011;2:11–13.
6. de Castro SMM, Ünlü Ç., Steller EP, *et al*. Evaluation of the Appendicitis Inflammatory Response score for patients with acute appendicitis. *World J Surg* 2012;36:1540–5.
7. Von-Muhlen B, Franzon O, BEDUSCHI MG, *et al*. Air score assessment for acute appendicitis. *ABCD* 2015;28:171–3.
8. Kollár D, McCartan DP, Bourke M, *et al*. Predicting acute appendicitis? A comparison of the Alvarado score, the Appendicitis



- Inflammatory Response score and clinical assessment. *World J Surg* 2015;39:104–9.
9. Ohle R, O'Reilly F, O'Brien KK, *et al*. The Alvarado score for predicting acute appendicitis: a systematic review. *BMC Medicine* 2011;9.
  10. Cochon L, Peña M, Baez A. Comparative Assessment of Diagnostic Accuracy of Computed Tomography and Ultrasonography in the Diagnosis of Acute Appendicitis After the Application of the Alvarado Scoring System. *Ann Emerg Med* 2013;62.
  11. Beardsley C, Dillon A, Nguyen F, *et al*. And paediatric appendicitis: a story of great expectations and the phantom appendix. *Arch Dis Child* 2016;99:A1–A620.
  12. Schneider C, Kharbanda A, Bachur R. Evaluating appendicitis scoring systems using a prospective pediatric cohort. *Ann Emerg Med* 2007;49:778–84.
  13. Scott AJ, Mason SE, Arunakirinathan M, *et al*. Risk stratification by the Appendicitis Inflammatory Response score to guide decision-making in patients with suspected appendicitis. *Br J Surg* 2015;102:563–72.
  14. Toprak H, Kilincaslan H, Ahmad IC, *et al*. Integration of ultrasound findings with Alvarado score in children with suspected appendicitis. *Pediatr Int* 2014;56:95–9.
  15. Malyar A, Singh B, Dar H, *et al*. A comparative study of Appendicitis Inflammatory response (air) score with Alvarado score in diagnosis of acute appendicitis. *Balkan Military Medical Review* 2015;18:72–6.
  16. Andersson M, Andersson RE. The Appendicitis Inflammatory Response score: a tool for the diagnosis of acute appendicitis that outperforms the Alvarado score. *World J Surg* 2008;32:1843–9.
  17. Pershad J, Waters TM, Langham MR, *et al*. Cost-effectiveness of diagnostic approaches to suspected appendicitis in children. *J Am Coll Surg* 2015;220:738–46.
  18. Sammalkorpi HE, Mentula P, Leppäniemi A. A new adult appendicitis score improves diagnostic accuracy of acute appendicitis—a prospective study. *BMC Gastroenterol* 2014;14.
  19. Blomqvist PG, Andersson REB, Granath F, *et al*. Mortality after appendectomy in Sweden, 1987–1996. *Ann Surg* 2001;233:455–60.
  20. York D, Smith A, Phillips JD, *et al*. The influence of advanced radiographic imaging on the treatment of pediatric appendicitis. *J Pediatr Surg* 2005;40:1908–11.