

Comparison of Predictive Validity of Alvarado Score and Appendicitis Inflammatory Response (AIR) Score, A Hospital Based Observational Study

Madasi V.

Dr. Venkaiah. Madasi, Assistant professor, Department of General Surgery, Rajiv Gandhi Institute of medical Sciences (RIMS), Ongole, Andhra Pradesh.

Address for Correspondence: Dr. Venkaiah. Madasi, Assistant Professor, Department of General Surgery, Rajiv Gandhi Institute of Medical Sciences (RIMS), Ongole, Andhra Pradesh. E-mail: drmadasi@gmail.com

Abstract

Introduction: Various new risk stratification scores have been proposed to accurately diagnose appendicitis. In the wake of limited number of comparative studies of these new scores, with the existing scores, the current study has compared the validity and reliability of Alvarado score and AIR score in diagnosis of appendicitis in a tertiary care teaching hospital. **Materials & Methods:** The current study was a prospective observational study. conducted in a tertiary care teaching hospital in south India, between July 2015 to August 2016, for a 12-month period. A total of 297 eligible subjects were included. For each patient Alvarado score and AIR score were calculated and compared with histopathological evaluation. **Results:** The predictive validity of Alvarado score as assessed by area under the ROC curve was 0.74 (95% CI 0.62 to 0.85), as compared to 0.95 (95 % CI 0.92 to 0.98) for AIR score. The sensitivity of the AIR score was 95.7% as compared to 87.3% sensitivity of ALVARADO score. AIR score had specificity of 90.5%, as compared to 52.4% for Alvarado score. Correspondingly, both false positive (47.6% vs. 9.5%) and false negative (12.7% vs.4.3%) rates were higher for Alvarado score. The positive and negative predictive values of Alvarado score were 96% and 23.9%, as compared to 99.2% and 61.3% for AIR score. The overall diagnostic accuracy of Alvarado score was 85%, as compared to 95% for AIR score. **Conclusions:** The newly proposed appendicitis inflammatory response score had displayed a better validity and reliability, as compared to Alvarado score

Keywords: Appendicitis, Alvarado score, AIR score, Predictive validity

Introduction

Appendicitis even though one of the most commonly treated condition by surgical interventions, can still pose a diagnostic dilemma to the surgeon [1]. There are many studies in the past, which have reported various proportions of negative appendectomy rates [2].

The negative appendectomy rates have been reported to come down drastically with the introduction of ultrasonography initially and Computerized tomography (CT) later [3-6].

But in resource poor settings there is still heavy reliance on clinical judgment as availability and quality of ultrasonography is quite variable. Performing routine CT may not be advisable and feasible in these settings considering the availability, cost and risk of radiation [1, 7].

Instead of subjective clinical judgment, various risk stratification score have been proposed to accurately diagnose appendicitis [8-12].

The Alvarado score which was proposed in the year 1986 has been one of the most widely used and evaluated

Manuscript Received: 24th September 2016

Reviewed: 5th October 2016

Author Corrected: 16th October 2016

Accepted for Publication: 30th October 2016

scoring system [8]. Various new scores have been proposed in recent times, which has claimed better validity and reliability [11, 13, 14].

Appendicitis inflammatory response (AIR) score is one such score, proposed by Anderson, M et al in 2008 [9] which has claimed much superior performance as compared to Alvarado score [3, 15-17].

The studies comparing the two scores are limited on Indian subjects, hence the current study is planned with an objective of comparing the validity and reliability of Alvarado score and AIR score in diagnosis of appendicitis in a tertiary care teaching hospital

Materials & Methods

Study design: The current study was a prospective observational study

Study setting: The study was conducted in a tertiary care teaching hospital in south India,

Study duration: The data collection for the study was done between July 2015 to August 2016, for a 12 months period.

Study population: The study population included all the subjects presenting to the emergency department, with symptoms suggestive of acute appendicitis and underwent appendectomy after necessary evaluation.

Inclusion & exclusion criteria: The inclusion criteria of the study were people aged above 15 years, belonging to both the genders. Patients whose condition was critical and subjects with past history of appendectomy were excluded from the study.

Sample size and sampling method: The study had included all the 297 eligible patients, who satisfied the inclusion criteria and were willing to provide informed

written consent were included in the study, hence no sampling was done.

Ethical issues: The study was approved by institutional human ethics committee. Informed written consent was obtained from all the study participants, after explaining the risks and benefits involved in the study. Confidentiality of the study participants was maintained throughout the study.

Study procedure: All the eligible subjects were evaluated by clinical examination, appropriate laboratory investigations and ultrasonography. For each patient Alvarado score [8] and AIR score [9] were calculated. Ultrasonography of the abdomen was performed on each subject.

Patient who were diagnosed as definitive case of acute appendicitis, as per the institute's protocol were taken to the surgery for open or laparoscopic appendectomy. The excised specimens of appendix were subjected to histopathological evaluation.

Statistical analysis: The data was summarized by mean and standard deviation for quantitative variables, frequency and proportion for categorical variables.

Patients were categorized as high or low risk as per the suggested cut off values of the two risk scoring systems.

The association between the scores and the HPE findings was assessed by cross tabulation and chi square test. Predictive validity of the Alvarado score and AIR score was assessed by ROC analysis. Area under the ROC curve along with its 95% CI and P- value were presented.

The sensitivity, specificity, predictive values and diagnostic accuracy of both the risk stratifications cores against HPE findings (Gold standard) were calculated and compared. IBM SPSS statistical software version 22 was used for statistical analysis [18].

Results

A total of 297 subjects were included in the final analysis. Majority of the study subjects belonged to 21 to 0 years of age.

Table-1: Age and gender distribution and test results in study population (N=297).

Parameter	Frequency	Percent
I. Age Group		
20 or below	60	20.20%
21 to 40	181	60.90%
41 to 60	43	14.50%
61 and above	13	4.40%
II. Sex		
Female	134	45.10%
Male	163	54.90%
III. Alvarado score		
High	251	84.50%
Low	46	15.50%
IV. AIR score		
High	266	89.60%
Low	31	10.40%
V. HPE		
Positive	276	92.90%
Negative	21	7.10%

The proportion of males and females was 54.9% and 45.1% respectively. The number of subjects stratified as high risk by Alvarado score were 251(84.50%0. AIR score has classified 266 (89.60%) subjects as high risk and 276 (92.90%) subjects were confirmed as appendicitis by HPE. (Table 1)

There was a statistically significant association between the Alvarado score, AIR score categories and HPE diagnosis of appendicitis. (Table 2)

Table-2: Association between the risk scores and HPE findings in study population.

Parameter	HPE		Chi-Square Value	P value
	Positive (N=276)	Negative (N=21)		
Alvarado score				
High	241 (87.30%)	10 (47.60%)	23.498a	<0.001
Low	35 (12.70%)	11(52.40%)		
AIR Score				
High	264 (95.70%)	2 (9.50%)	154.858a	<0.001
Low	12 (4.30%)	19 (90.50%)		

The predictive validity of Alvarado score as assessed by area under the ROC curve was 0.74 (95% CI 0.62 to 0.85), as compared to 0.95 (95 % CI 0.92 to 0.98) for AIR score. (Table 3 and figure 1)

Table-3: Comparison ROC analysis parameters of both risk scores in study population.

Risk Score	Area Under the curve (AUC)	Asymptotic 95% Confidence Interval		P value
		Lower Bound	Upper Bound	
Alvarado Score	0.74	0.62	0.85	< 0.001
AIR score	0.95	0.92	0.98	< 0.001

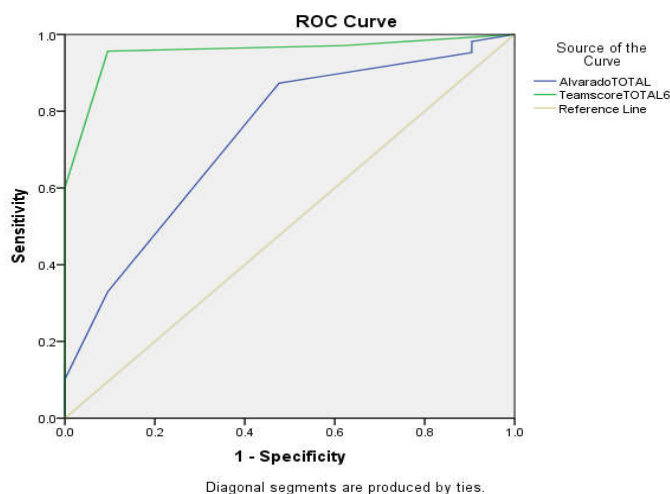


Figure-1: ROC analysis to assess the predictive validity of Alvarado and AIR scores

The sensitivity of the AIR score was 95.7% as compared to 87.3% sensitivity of ALVARADO score. AIR score had a specificity of 90.5%, as compared to 52.4% for Alvarado score. Correspondingly, both false positive (47.6% vs. 9.5%) and false negative (12.7% vs. 4.3%) rates were higher for Alvarado score. The positive and negative predictive values of Alvarado score were 96% and 23.9%, as compared to 99.2% and 61.3% for AIR score. The overall diagnostic accuracy of Alvarado score was 85%, as compared to 95% for AIR score. (Table 4)

Table-4: Comparison of validity, predictive values and reliability of the two risk scores.

Parameter	Alvarado score Parameter (95% CI)	AIR score Parameter (95% CI)
Sensitivity	87.30% (83.39% to 91.24%)	95.70% (93.24% to 98.05%)
Specificity	52.40% (31.02% to 73.74%)	90.50% (77.92% to 103.0%)
False positive rate	47.60% (26.25% to 68.97%)	9.50% (-3.03% to 22.07%)
False negative rate	12.70% (8.755% to 16.60%)	4.30% (1.941% to 6.753%)
Positive predictive value	96.00% (93.59% to 98.43%)	99.20% (98.21% to 100.2%)
Negative predictive value	23.90% (11.58% to 36.23%)	61.30% (44.14% to 78.43%)
Diagnostic accuracy	85% (80.77% to 88.92%)	95% (92.87% to 97.69%)

The reliability of the risk scores, as measured by kappa statistic was considerably higher for AIR score (0.706), compared to Alvarado score (0.256), which was statistically significant (P value < 0.001). (Table 5)

Table-5: Comparison of reliability of the two risk scores in study population.

Risk Score	Kappa statistic	Standard error	P value
Alvarado Score	0.256	0.077	< 0.001
AIR score	0.706	0.074	< 0.001

Discussion

Considering the non-availability of advanced investigations like CT, risk stratification scores are valuable tools in reducing diagnostic dilemma in acute appendicitis in resource poor settings [1]. But concern regarding poor validity and reliability and the resulting negative appendectomy rates, have prevented their widespread use in clinical practice [16].

With the advent of many new scoring systems, which have claimed superiority over existing scores, it is imperative to test this claim in different population subgroups before recommending their use in routine practice [10, 11, 19-22]. The current study has compared the validity and reliability of newly introduced AIR score with Alvarado score. In the current study, the predictive validity of Alvarado score as assessed by area under the ROC curve was 0.74 (95% CI 0.62 to 0.85), as compared to 0.95 (95 % CI 0.92 to 0.98) for AIR score.

The reliability of the risk scores, as measured by kappa statistic was considerably higher for AIR score (0.706), compared to Alvarado score (0.256), which was statistically significant (P value < 0.001). De Castro, S. M., et al. [15] have reported an AUC of 0.96 for AIR score and 0.82 for Alvarado score (p < 0.05). Macco, S., et al [12] have reported an area under the receiver-operating curve of 0.90 for AIR score and 0.87 for Alvarado score was 0.87.

Andersson, M. and R. E. Andersson[9], who, while proposing the AIR score have reported an ROC area of the 0.97 for advanced appendicitis and 0.93 for all appendicitis. Alvarado score had an ROC area of 0.92 and 0.88 respectively for advanced and all appendicitis.

Sensitivity of the AIR score was 95.7% as compared to 87.3% sensitivity of ALVARADO score. AIR score had s

specificity of 90.5%, as compared to 52.4% for Alvarado score. Correspondingly, both false positive (47.6% vs. 9.5%) and false negative (12.7% vs.4.3%) rates were higher for Alvarado score. The positive and negative predictive values of Alvarado score were 96% and 23.9%, as compared to 99.2% and 61.3% for AIR score. The overall diagnostic accuracy of Alvarado score was 85%, as compared to 95% for AIR score. In study by Macco, S., et al [12].

AIR has shown better specificity and positive predictive value than that of the Alvarado score. In study by Andersson, M. and R. E. Andersson [9] “Sixty-three percent of the patients were classified into the low- or high-probability group with an accuracy of 97.2%, leaving 37% for further investigation.

Seventy-three percent of the nonappendicitis patients, 67% of the advanced appendicitis, and 37% of all appendicitis patients were correctly classified into the low- and high-probability zone, respectively.” De Castro, S. M., et al. [15] the AIR score was reported to outperform the Alvarado score in diagnosis of appendicitis in difficult patient groups like women, children, and the elderly.

Kollar, D., et al [3] have reported substantially higher specificity (97 %) and positive predictive value (88 %) for AIR score. As compared to than the Alvarado score (76 and 65 %, respectively).

Conclusions

1. The newly proposed appendicitis inflammatory response score had displayed a better validity and reliability, as compared to Alvarado score.
2. Both negative appendectomy rates and missing cases of appendicitis will be reduced, if AIR score is used for treatment decisions, in place of Alvarado score

Recommendations

1. Further large scale studies in different settings and different population groups are necessary to further strengthen the evidence in this regard
2. The variations in validity and reliability in specific population subgroups, like females, pediatric population, obese people, elderly etc. have to be studied

Acknowledgements: We would like to acknowledge the statistical support provided by Dr. Murali Mohan Reddy of Evidencian research Foundation.

Conflict of interest: None declared.

Funding: Nil, **Permission from IRB:** Yes

References

1. Shogilev DJ, Duus N, Odom SR, Shapiro NI. Diagnosing appendicitis: evidence-based review of the diagnostic approach in 2014. *West J Emerg Med.* 2014 Nov; 15(7): 859-71. doi: 10.5811/westjem. 2014.9. 21568. Epub 2014 Oct 7.
2. Mariadason JG, Wang WN, Wallack MK, Belmonte A, Matari H. Negative appendectomy rate as a quality metric in the management of appendicitis: impact of computed tomography, Alvarado score and the definition of negative appendectomy. *Ann R Coll Surg Engl.* 2012 Sep; 94 (6): 395 - 401. doi: 10.1308 / 003588412 X 13171221592131.
3. Kollár D, McCartan DP, Bourke M, Cross KS, Dowdall J. Predicting acute appendicitis? A comparison of the Alvarado score, the Appendicitis Inflammatory Response Score and clinical assessment. *World J Surg.* 2015 Jan;39(1):104-9. doi: 10.1007/s00268-014-2794-6.
4. Liu W, Wei Qiang J, Xun Sun R. Comparison of multislice computed tomography and clinical scores for diagnosing acute appendicitis. *J Int Med Res.* 2015 Jun;43(3):341-9. doi: 10.1177/0300060514564475. Epub 2015 Mar 11.
5. Apisarntharak P, Suvannarerg V, Pattaranutaporn P, Charoensak A, Raman SS, Apisarntharak A. Alvarado score: can it reduce unnecessary CT scans for evaluation of acute appendicitis? *Am J Emerg Med.* 2015

Feb;33(2):266-70. doi: 10.1016/j.ajem.2014.11.056. Epub 2014 Dec 3.

6. Cochon L, Esin J, Baez AA. Bayesian comparative model of CT scan and ultrasonography in the assessment of acute appendicitis: results from the Acute Care Diagnostic Collaboration project. *Am J Emerg Med.* 2016 Nov;34(11):2070-2073. doi: 10.1016/j.ajem.2016.07.012. Epub 2016 Jul 16.

7. Alvarado A. How to improve the clinical diagnosis of acute appendicitis in resource limited settings. *World J Emerg Surg.* 2016 Apr 26;11:16. doi: 10.1186/s13017-016-0071-8. eCollection 2016.

8. Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med.* 1986 May; 15(5):557-64.

9. Andersson M, Andersson RE. The appendicitis inflammatory response score: a tool for the diagnosis of acute appendicitis that outperforms the Alvarado score. *World journal of surgery.* 2008;32(8):1843-9.

10. Erdem H, Cetinkunar S, Das K, Reyhan E, Deger C, Aziret M, et al. Alvarado, Eskelinen, Ohmann and Raja Isteri Pengiran Anak Saleha Appendicitis scores for diagnosis of acute appendicitis. *World journal of gastroenterology.* 2013;19(47):9057-62.

11. González Del Castillo J, Ayuso FJ, Trenchs V, Martínez Ortiz de Zarate M, Navarro C, Altali K, Fernandez C, Huckins D, Martín-Sánchez FJ; representing INFURG-SEMES group. Diagnostic accuracy of the APPY1 Test in patients aged 2-20 years with suspected acute appendicitis presenting to emergency departments. *Emerg Med J.* 2016 Dec;33(12):853-859. doi: 10. 1136/emermed-2015-205259. Epub 2016 Sep 9.

12. Macco S, Vrouenraets BC, de Castro SM. Evaluation of scoring systems in predicting acute appendicitis in children. *Surgery.* 2016 Dec;160(6):1599-1604. doi: 10. 1016/j. surg.2016.06.023. Epub 2016 Aug 12.

13. Khanafer I, Martin DA, Mitra TP, Eccles R, Brindle ME, Nettel-Aguirre A, Thompson GC. Test characteristics of common appendicitis scores with and without laboratory investigations: a prospective observational

study. *BMC Pediatr.* 2016 Aug 30;16(1): 147. doi: 10.1186/s12887-016-0687-6.

14. N N, Mohammed A, Shanbhag V, Ashfaq K, S A P. A Comparative Study of RIPASA Score and ALVARADO Score in the Diagnosis of Acute Appendicitis. *J Clin Diagn Res.* 2014 Nov;8(11):NC03-5. doi: 10.7860/JCDR/2014/9055.5170. Epub 2014 Nov 20.

15. de Castro SM, Unlu C, Steller EP, van Wagenveld BA, Vrouwenraets BC. Evaluation of the appendicitis inflammatory response score for patients with acute appendicitis. *World journal of surgery.* 2012;36(7): 1540-5.

16. Mán E, Simonka Z, Varga A, Rárosi F, Lázár G. Impact of the Alvarado score on the diagnosis of acute appendicitis: comparing clinical judgment, Alvarado score, and a new modified score in suspected appendicitis: a prospective, randomized clinical trial. *Surg Endosc.* 2014 Aug; 28(8):2398-405. doi: 10.1007/s00464-014-3488-8. Epub 2014 Apr 5.

17. Pogorelić Z, Rak S, Mrklič I, Jurić I. Prospective validation of Alvarado score and Pediatric Appendicitis Score for the diagnosis of acute appendicitis in children.

Pediatr Emerg Care. 2015 Mar;31(3):164-8. doi: 10.1097/PEC.0000000000000375.

18. SPSS I. IBM SPSS statistics 22. Algorithms Chicago: IBM SPSS Inc. 2013.

19. Di Saverio S, Birindelli A, Piccinini A, Catena F, Biscardi A, Tugnoli G. How Reliable Is Alvarado Score and Its Subgroups in Ruling Out Acute Appendicitis and Suggesting the Opportunity of Nonoperative Management or Surgery? *Annals of surgery.* 2016.

20. Jalil A, Shah SA, Saaiq M, Zubair M, Riaz U, Habib Y. Alvarado scoring system in prediction of acute appendicitis. *J Coll Physicians Surg Pak.* 2011 Dec;21(12):753-5. doi: 12.2011/JCPS.753755.

21. Kariman H, Shojaee M, Sabzghabaei A, Khatamian R, Derakhshanfar H, Hatamabadi H. Evaluation of the Alvarado score in acute abdominal pain. *Ulusal travma ve acil cerrahi dergisi = Turkish journal of trauma & emergency surgery : TJTES.* 2014;20(2):86-90.

22. Sousa-Rodrigues CF, Rocha AC, Rodrigues AK, Barbosa FT, Ramos FW, Valoes SH. Correlation between the Alvarado Scale and the macroscopic aspect of the appendix in patients with appendicitis. *Revista do Colegio Brasileiro de Cirurgioes.* 2014;41(5):336-9.

.....
How to cite this article?

Madasi V. Comparison of Predictive Validity of Alvarado Score and Appendicitis Inflammatory Response (AIR) Score, A Hospital Based Observational Study. *Int J surg Orthopedics* 2016;2(3):30-36.doi: 10.17511/ijoso.2016.i03.02
.....