



# Prospective Comparison of the Efficacy of Two Common Appendicitis Scoring Systems: Is Combination a Solution?

## İki Apandisit Skorlama Sisteminin Etkinliğinin Prospektif Olarak Karşılaştırılması: Kombinasyon Bir Çözüm mü?

© Mehmet Üstün<sup>1</sup>, © Avni Can Karaca<sup>2</sup>, © Semra Demirli Atıcı<sup>1</sup>, © Göksever Akpınar<sup>1</sup>, © Cem Karaali<sup>1</sup>

<sup>1</sup>University of Health Sciences Turkey, Tepecik Training and Research Hospital, Clinic of General Surgery, İzmir, Turkey

<sup>2</sup>İzmir University of Economics Faculty of Medicine, Department of General Surgery, İzmir, Turkey

### ABSTRACT

**Aim:** The diagnosis of acute appendicitis mostly relies on history taking and physical examination findings supported by laboratory and imaging studies. A number of different diagnostic scoring systems have been developed to facilitate diagnosis, and their accuracies vary among patient populations. This prospective study aims to evaluate the accuracy of the two most frequently used scoring systems in the Turkish patient population and to analyse the possible diagnostic advantage of using these two systems in combination.

**Method:** Patients admitted to the emergency department of a tertiary healthcare centre with acute abdominal pain who eventually underwent appendectomy between July 2018 and January 2019 were enrolled in the study. Alvarado and Raja Isteri Pengiran Anak Saleha Appendicitis (RIPASA) scores, as well as other laboratory parameters, were recorded for each patient. Using histopathologic examination as the gold standard, the sensitivity, specificity and positive and negative predictive values of each scoring system were calculated and combined using McNemar's x2 test.

**Results:** Data from a total of 203 patients were analysed. The sensitivity of the RIPASA system (95%) was far superior to that of the Alvarado system (35.6%). However, the Alvarado scoring system had much higher diagnostic specificity than the RIPASA system (80% vs 33.3%). The combined sensitivity and specificity of the tests rose to 88% and 62.5%, respectively.

**Conclusion:** The RIPASA system has high sensitivity; however, the Alvarado system has high specificity for the Turkish population. Both the Alvarado and RIPASA scoring systems are useful clinical tools with different strengths. Using these two systems in combination increases diagnostic power by combining the strongest aspects of both tests.

**Keywords:** Alvarado, RIPASA, appendicitis, diagnosis

### ÖZ

**Amaç:** Akut apandisit tanısı çoğunlukla laboratuvar ve görüntüleme çalışmalarıyla desteklenen anamnez ve fizik muayene bulgularına dayanır. Teşhisi kolaylaştırmak için bir dizi farklı teşhis skorlama sistemi geliştirilmiştir ve bunların doğrulukları hasta popülasyonları arasında değişmektedir. Bu prospektif çalışma, Türk hasta popülasyonunda en sık kullanılan iki skorlama sisteminin doğruluğunu değerlendirmeyi ve bu iki sistemi birlikte kullanmanın olası tanılma avantajını incelemeyi amaçlamaktadır.

**Yöntem:** Temmuz 2018 ile Ocak 2019 tarihleri arasında, 3. basamak sağlık hizmeti veren kuruluşun acil servisine akut karın ağrısı ile başvuran ve ardından apandisit ameliyatı geçiren hastalar çalışmaya alındı. Her hasta için Alvarado ve Raja Isteri Pengiran Anak Saleha Apandisit (RIPASA) skorları ve diğer laboratuvar parametreleri kaydedildi. Histopatolojik inceleme altın standart tanı yöntemi olarak kullanılarak, her puanlama sisteminin duyarlılığı, özgüllüğü ve pozitif-negatif prediktif değerleri McNemar'ın x2 testi kullanılarak hesaplandı.

**Bulgular:** Toplam 203 hastanın verileri analiz edildi. RIPASA sisteminin duyarlılığı (%95) Alvarado sistemininkinden (%35,6) çok daha üstündü. Ancak Alvarado skorlama sistemi, RIPASA sisteminden çok daha yüksek tanılma özgüllüğe sahipti (%80'e karşı %33,3). Testlerin birleşik duyarlılığı ve özgüllüğü sırasıyla %88 ve %62,5'e yükseldi.

**Sonuç:** RIPASA sistemi yüksek hassasiyete sahiptir; ancak Alvarado sistemi Türk nüfusu için yüksek özgüllüğe sahiptir. Alvarado ve RIPASA skorlama sistemleri, farklı güçlü yönleri sahip faydalı klinik araçlardır. Bu iki sistemi birlikte kullanmak, her iki testin en güçlü yönlerini birleştirerek teşhis gücünü artıracaktır.

**Anahtar Kelimeler:** Alvarado, RIPASA, apandisit, teşhis



Address for Correspondence/Yazışma Adresi: Mehmet Üstün MD,

University of Health Sciences Turkey, Tepecik Training and Research Hospital, Clinic of General Surgery, İstanbul, Turkey

E-mail: dr.m.ustun@gmail.com ORCID ID: orcid.org/0000-0003-2646-5239

Received/Geliş Tarihi: 10.01.2020 Accepted/Kabul Tarihi: 02.04.2020

## Introduction

Acute appendicitis (AA) remains one of the most common cause of abdominal emergencies, with a lifetime risk of approximately 7%.<sup>1</sup> The diagnosis of AA mostly relies on history taking and physical examination findings supported by laboratory and imaging studies. Not surprisingly, a number of different scoring systems facilitating the diagnosis of AA are suggested in the literature. Among them, the scoring system created by Alvarado in 1986<sup>2</sup> and later modified is one of the most accepted and commonly used around the world. However, an article by Chong et al.<sup>3</sup> in 2010 emphasised the low sensitivity and specificity of the Alvarado scoring system in Asian and Middle Eastern populations and suggested a substitute. The Raja Isteri Pengiran Anak Saleha Appendicitis (RIPASA) scoring system was then quickly adopted, and a number of validation studies, especially from Asia and the Middle East, followed its publication.<sup>4</sup> However, to the best of our knowledge, there are only three studies evaluating the efficacy of the RIPASA scoring system from Turkey, the largest covering 113 patients.<sup>5,6,7</sup>

With this prospective study, we aimed to compare the efficacy of these commonly used scoring systems on a larger patient

population, widening the projection of validation, and to propose different applications of both scoring systems.

## Materials and Methods

Patients admitted to the emergency department of a tertiary healthcare centre with acute abdominal pain who eventually underwent appendectomy between July 2018 and January 2019 were enrolled in the study. The exclusion criteria were pregnancy and refusal to consent. Regardless of the initial differential diagnosis, demographics, as well as Body Mass index (BMI), findings from physical examination and imaging studies and laboratory results were recorded, and Alvarado and RIPASA scores were calculated for each patient, using the scale charts given in Table 1.

Attending surgeons who carried out surgeries were blinded to the patients' scores; hence, all operation indications were established on the basis of findings of physical examinations and laboratory results. Data from patients with pathologies other than appendicitis on histopathologic examination were omitted. In concordance with the current literature, patients with Alvarado scores equal to and higher than 7 and RIPASA scores equal to or higher than 7.5 were classified as having "clinical appendicitis".

**Table 1.** Components and points used in the diagnostic scoring systems

Alvarado scoring system		RIPASA scoring system	
Component	Points	Component	Points
Migrating pain	1	Male sex	1
Anorexia	1	Female sex	0.5
Nausea or vomiting	1	Age <40 years	1
Tenderness in RLQ	2	Age >40 years	0.5
Rebound tenderness	1	Foreigner	1
Elevated body temperature (>38 °C)	1	Pain in RLQ	0.5
Leucocytosis (>10,000/mm <sup>3</sup> )	2	Nausea or vomiting	1
Neutrophilia (>70%)	1	Migrating pain	0.5
		Anorexia	1
		Symptom duration <48 hours	1
		Symptom duration >48 hours	0.5
		Hypersensitivity in RLQ	1
		Guarding	2
		Rebound tenderness	1
		Positive Rovsing's sign	2
		Fever	1
		Leucocytosis	1
		Negative urine findings	1

RLQ: Right lower quadrant

Using these patients' histopathologic examination results as the gold standard, the sensitivity, specificity, positive and negative likelihood ratios and predictive values of each diagnostic test were calculated using McNemar's  $\chi^2$  test. Confidence intervals for sensitivity and specificity were calculated as exact Clopper-Pearson confidence intervals. Confidence intervals for likelihood ratios were calculated using the "Log method", and predictive values were taken as the standard logit confidence intervals.<sup>8,9</sup> Approval from the institutional research ethics board was obtained (decision number 2018/8-1).

## Results

A total of 203 patients were enrolled in the study. There were slightly more female patients (n=104) than male patients (n=99). The mean patient age was 36.4 (range: 18-78, standard deviation: 14.15). The calculated average score was 6.75 (range: 3-9) for the Alvarado and 9.84 (range: 5-16.5) for the RIPASA scale. The average BMI of the patient group was 26, ranging from 17.7 to 49.3. Open surgery was the procedure of choice, with 83.3% (n=169) of the patients undergoing laparotomy and 16.7% (n=34) undergoing laparoscopy. The negative appendectomy rate was 7.4% (n=15) of the 203 appendectomies performed. Computer tomography (CT) was utilised as a diagnostic test in the majority of patients (82.8% n=168), and imaging findings were coherent with AA in 154 (91%). Of these 168 CT studies, there were 10 false-positive and 10 false-negative evaluations, making the sensitivity of CT 93.5% and the specificity 28.5% in the diagnosis of appendicitis. The Alvarado score was 7 or higher in 34.5% (n=70) patients, suggesting a strong probability of appendicitis in these patients. There were three false-positive and 121 false-

negative predictions of AA when 7 points was used as the cut-off value for the Alvarado scoring system. The sensitivity of the Alvarado scoring system was 35.6%, and the specificity was 80%. The positive and negative likelihood ratios were 1.78 and 0.80, respectively. Table 2 shows contingency tables and details of these values in terms of the 95% confidence interval (CI) for the Alvarado scoring system. Among the 203 analysed patients, there were 189 patients (93%) with RIPASA scores of 7.5 or higher. There were 10 false-positive and 9 false-negative predictions of AA when 7.5 points was used as a cut-off value for the RIPASA scoring system. The sensitivity of the RIPASA scoring system was 95.2%, and the specificity was 33.3%. The positive and negative likelihood ratios were 1.43 and 0.14, respectively. Table 3 shows contingency tables and details of these values in terms of the 95% CI for the RIPASA scoring system. The effect of BMI on the sensitivity and specificity of the diagnostic scoring systems was analysed. Patients were classified as either overweight (BMI  $\geq 25$ ) or normal weight. The sensitivity and specificity of the Alvarado scoring system was changed to 31.5% and 100%, respectively, in normal weight patients. These values were calculated as 37.8% and 62.5% in overweight individuals. Similarly, the sensitivity and specificity of the RIPASA scoring system was affected by BMI. These values were 98.6% and 57.1% in normal weight patients; however, they decreased to 94% and 12.5% in overweight individuals. The combined accuracy of the two tests was also investigated. A subgroup of 84 patients in whom the predictions of the two scoring systems coincided (both of the tests agreed on the prediction) was created from the patient population, and the accuracy for this subgroup was also analysed. The combined sensitivity and specificity of the tests were 88.1%

**Table 2.** Contingency table and details for the statistical accuracy of the Alvarado scoring system

Alvarado Score	Pathologic examination		Total
	Not appendicitis	Appendicitis	
<7	12	121	133
7	3	67	70
	Value	95% CI	
Sensitivity	35.64%	28.80%-42.93%	
Specificity	80.00 %	51.91%-95.67%	
Positive LR	1.78	0.64-4.99	
Negative LR	0.80	0.61-1.06	
Positive PV	95.71%	88.85%-98.43%	
Negative PV	9.02%	7.01%-11.54%	
Accuracy	38.92%	32.17%-45.99%	

LR: Like hood ratio, PV: Predictive value, CI: Confidence interval

and 62.5%, respectively. Details of this subgroup of patients are summarised in Table 4.

## Discussion

As AA is a very common cause of abdominal emergencies, its diagnosis has been studied frequently, and the literature contains multiple validation and comparison studies of the two most commonly utilised diagnostic scoring systems: the Alvarado and the RIPASA systems. A recent meta-analysis on the subject clearly shows that, although it varies among studies, the sensitivity of the Alvarado scoring system is consistently lower than that of the RIPASA scoring system.<sup>4</sup> This phenomenon was also validated in the current study: the calculated sensitivity of the Alvarado scoring system was markedly lower (35.6%) than the 95% sensitivity rate of

the RIPASA system. Similar sensitivity rates have also been reported from different Turkish patient populations.<sup>5</sup>

In contrast, when the specificities of the two tests were compared, the Alvarado scoring system seemed to be much more precise than the RIPASA system for diagnosing AA (80% vs 33.3% specificity). With a few exceptions, the literature is also in agreement on the lower specificity provided by the RIPASA system.<sup>4</sup>

CT imaging has been used frequently (82.8%) as a diagnostic tool in this patient population. A retrospective reanalysis of the data revealed that the majority of CT imaging studies had been ordered by the attending physician in the emergency department before the patient had been seen by a specialist. The frequent utilisation of CT imaging could also be attributed to the study design, since the patients were

**Table 3.** Contingency table and details for the statistical accuracy of the RIPASA scoring system

RIPASA score	Pathologic examination		Total
	Not appendicitis	Appendicitis	
<7.5	5	9	14
7.5	10	179	189
	Value	95% CI	
Sensitivity	95.21%	91.11%-97.79%	
Specificity	33.33%	11.82%-61.62%	
Positive LR	1.43	1.00-2.05	
Negative LR	0.14	0.06-0.37	
Positive PV	94.71%	92.59%-96.25%	
Negative PV	35.71%	17.56%-59.16%	
Accuracy	90.64%	85.77%-94.27%	

LR: Like hood ratio, PV: Predictive value, CI: Confidence interval

**Table 4.** Contingency table and details for the statistical accuracy of the scoring systems combined

Combined Prediction	Pathologic examination		Total
	Not appendicitis	Appendicitis	
Negative	5	9	14
Positive	3	67	70
	Value	95% CI	
Sensitivity	88.16%	78.71%-94.44%	
Specificity	62.50%	24.49%-91.48%	
Positive LR	2.35	0.96-5.77	
Negative LR	0.19	0.08-0.43	
Positive PV	95.71%	90.09%-98.21%	
Negative PV	35.71%	19.74%-55.66%	
Accuracy	85.71%	76.38%-92.39%	

LR: Like hood ratio, PV: Predictive value, CI: Confidence interval

immediately included in the study as they were admitted to the emergency department with findings of acute abdominal pain before a diagnosis of AA had been established. Nonetheless, the 93.5% sensitivity of CT imaging clearly has no apparent diagnostic advantage over the 95% sensitivity provided by the RIPASA scoring system. This being said, the influence of BMI on the sensitivity of these diagnostic tests should always be kept in mind, and imaging studies are potentially helpful in those situations.

In addition, oversensitivity of a diagnostic tool is not always a desired outcome, since it can eventually lead to an increased number of unnecessary appendectomies. In fact, the literature advocates keeping negative appendectomy rates lower than 15% but simultaneously reducing the incidence of late diagnoses.<sup>10</sup> The RIPASA system has high sensitivity (95%); however, the Alvarado system has high specificity (80%). Therefore, using these two scoring systems in combination can pose an alternative solution by harnessing the synergistic diagnostic power of the successful aspects of both tests together. Indeed, when a subgroup of 84 patients with exact coinciding predictions on both tests was analysed, the combined sensitivity and specificity of the tests came to 88% and 62.5%, respectively. This rate of sensitivity covers the suggested limit rate of 15% negative appendectomies found in the literature.

### Study Limitations

The limitation of this study is that it was performed in a single centre, and the number of patients was insufficient to reflect the characteristics of a society.

### Conclusion

Both the Alvarado and RIPASA scoring systems are useful clinical tools with different strengths. Using these two systems in combination increases diagnostic power by combining strongest aspects of both tests.

### Ethics

**Ethics Committee Approval:** Approval from the institutional research ethics board was obtained (decision number 2018/8-1).

**Informed Consent:** Written informed consent was obtained from each patient who participated in this study.

**Peer-review:** Internally and externally peer reviewed.

### Authorship Contributions

Surgical and Medical Practices: M.Ü., A.C.K., S.D.A., G.A., C.K., Concept: M.Ü., A.C.K., Design: M.Ü., A.C.K., Data Collection or Processing: S.D.A., G.A., Analysis or Interpretation: M.Ü., A.C.K., C.K., Literature Search: M.Ü., A.C.K., G.A., C.K., Writing: M.Ü.

**Conflict of Interest:** No conflict of interest was declared by the authors.

**Financial Disclosure:** The authors declared that this study received no financial support.

### References

1. Addiss DG, Shaffer N, Fowler BS, Tauxe RV. The epidemiology of appendicitis and appendectomy in the United States. *Am J Epidemiol* 1990;132:910-925.
2. Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med* 1986;15:557-564.
3. Chong CF, Adi MI, Thien A, Suyoi A, Mackie AJ, Tin AS, et al. Development of the RIPASA score: a new appendicitis scoring system for the diagnosis of acute appendicitis. *Singapore Med J* 2010;51:220-225.
4. Frountzas M, Stergios K, Kopsini D, Schizas D, Kontzoglou K, Toutouzias K. Alvarado or RIPASA score for diagnosis of acute appendicitis? A meta-analysis of randomized trials. *Int J Surg* 2018;56:307-314.
5. Ferlengöz E, Ferlengöz AG, Akbulut H, Kadioglu H. Akut Apendisit Değerlendirilmesinde Dört Farklı Skorum Sisteminin Değerlendirilmesi; İleri Dönük Klinik Çalışma. Evaluation of Four Different Scoring Systems in the Management of Acute Appendicitis; a Prospective Clinical Study 2013;51:15-17.
6. Unal Ozdemir Z, Ozdemir H, Sunamak O, Akyuz C, Torun M. Comparison of the reliability of scoring systems in the light of histopathological results in the diagnosis of acute appendicitis. *Hong Kong Journal of Emergency Medicine* 2019;26:323-327.
7. Erdem H, Cetinkunar S, Das K, Reyhan E, Deger C, Aziret M, et al. Alvarado, Eskelinen, Ohhmann and Raja Isteri Pengiran Anak Saleha Appendicitis scores for diagnosis of acute appendicitis. *World J Gastroenterol* 2013;19:9057-9062.
8. Mercaldo ND, Lau KF, Zhou XH. Confidence intervals for predictive values with an emphasis to case-control studies. *Stat Med* 2007;26:2170-183.
9. Altman DG. Diagnostic Tests. In: Douglas Altman DM, Trevor Bryant, Martin Gardner, eds. *Statistics with Confidence: Confidence Intervals and Statistical Guidelines*. 2 th ed: BMJ Books; 2000:105-119.
10. Flum DR, Morris A, Koepsell T, Dellinger EP. Has misdiagnosis of appendicitis decreased over time? A population-based analysis. *JAMA* 2001;286:1748-1753.