

Clinical Scoring System: A Valuable Tool for Decision Making in cases of Acute Appendicitis

Pages with reference to book, From 254 To 259

Ambreen Jawaid, Ansul Asad, Arashk Motiei, Asma Munir, Erum Bhutto, Haroon Choudry, Kamran Idrees, Meher Rahman, Mona Ahuja, Qurrat-ul-ain Nawab, Raheel Ahmed, Sadia Ali, Saima Aslam, Saleha Abbasi, Sharmeen Feerasta, Sonia Alam, Uzma Ahmed (Final Year Medical Students, The Aga Khan University, Karachi.)
Imtiaz Jehan (Senjor Instructor, Department of Community Health Sciences, The Aga Khan University, Karachi.)

Abstract

Objective: Decision making in cases of acute appendicitis poses a clinical challenge specially in developing countries where advanced radiological investigations do not appear cost effective and so clinical parameters remain the mainstay of diagnosis. The aim of our study was to devise a scoring system from our local database and test its accuracy in the preoperative diagnosis of acute appendicitis.

Methods: Clinical data from 401 patients having undergone appendectomy were collected to identify predictive factors that distinguished those with appendicitis from those who had a negative appendectomy. Ten such factors were identified and using Bayesian probability a weight was assigned to each and the results summated to get an overall score. A cut-off point was identified to separate patients for surgery and those for observation. The scoring system was then retrospectively applied to a second population of 99 patients in order to compare suggested actions (derived from the scoring system) to those actually taken by surgeons. The sensitivity, specificity and accuracy for the level of decision was then calculated.

Results: Of the 99 patients, the method suggested immediate surgery for 65 patients, 63 of whom had acute appendicitis (3.1% diagnostic error rate). Of the 33 patients in whom the score suggested active observation, 18 had appendicitis. The accuracy of our scoring system was 82%. The method had a sensitivity of 78%, specificity 89% and a positive predictive value of 97%. The negative appendectomy rate determined by our study was 7% and the perforation rate 13%.

Conclusion: Scoring system developed from a local database can work effectively in routine practice as an adjunct to surgical decision making in questionable cases of appendicitis (JPMA 49:254, 1999).

Introduction

Acute appendicitis is the most common cause leading to emergency abdominal surgery, accounting for 10-30% of acute abdominal conditions according to two studies from Pakistan^{1,2}. Although improvements in preoperative diagnosis have been made with the use of imaging techniques, the diagnosis still remains a challenge in developing countries where such technology is not freely available.

Diagnostic difficulties occur when patients present with atypical findings resulting in negative appendectomies. A negative appendectomy is taken as a surgery performed for a preoperative diagnosis of appendicitis that results in a normal histopathological specimen. Rates of negative appendectomies range between 8-35% with increased rates (up to 45%) seen in women in the reproductive age group³.

The negative appendectomy rate reported by Rehman et al from Abbotabad, Pakistan is 18%².

In the 1950s a 20-25% negative appendectomy rate was proposed as acceptable in order to minimize the incidence of perforated appendicitis and resulting high morbidity and mortality⁴. This implies that the rate of perforation is related to a delay in diagnosis and/or treatment, and that by accepting a higher negative appendectomy rate one can, in effect, buy a lower perforation rate⁵. However, recent studies

propose that the rate of perforation is due to a delay in patient presentation, rather than a delay in treatment⁶ suggesting that the incidence of negative appendectomies can be lowered without compromising the perforation rate. Negative appendectomy rates remained relatively stable over the last 70 years. However, with the introduction of CT scanning in developed countries in the last five years the rate of negative appendectomies has decreased from 16% to 4% in the general population and from 25-45% to 8% in female patients of childbearing age⁷. Furthermore the striking decrease in the negative appendectomy rate has been achieved without an increase in the perforation rate or mortality. The perforation rate in one series was 21%⁷ after the introduction of CT scanning compared to 21-23% in previously reported surgical surveys⁸⁻¹⁰. The sensitivity of CT scan in the diagnosis of acute appendicitis is reported to be 97% with a specificity of 97%⁷. In a third world country like Pakistan, availability and economic constraints limit the routine use of CT scan in patients with suspected appendicitis. In our setting acute appendicitis is diagnosed on the basis of clinical parameters. Different techniques have been devised to assist in equivocal cases in attempts to decrease negative appendectomy rates. Diagnostic scores are one such technique. These scores make use of history, physical examination and laboratory findings. Presently six scores have been proposed to aid the diagnosis of acute appendicitis¹¹⁻¹³. Although all authors have reported excellent predictive accuracy in their series, few have confirmed the reliability in subsequent studies¹²⁻¹³. The Alvarado score described in 1986 has subsequently been validated in adult surgical practice¹⁴. Its use in a prospective study of 215 adults and children decreased an unusually high negative appendectomy rate of 44% to 14%¹⁵. Ramirez et al 1994¹¹ created a new scoring system and tested its accuracy on a local database. The scoring system showed a sensitivity of 80% and a specificity of 81%. Their results confirm those of other authors¹³ and suggest that scoring system developed from a local database can become the ideal complimentary method in the diagnosis of suspected acute appendicitis. Scoring systems have not been used routinely in clinical practice in the Western world due to easy availability of CT scans and because of their high predictive accuracy only in the population on which they are devised. The aim of our study was to devise a scoring system based on our own local setting and to test its accuracy in the preoperative diagnosis of appendicitis.

Material and Methods

This study was conducted at the Aga Khan University Hospital in Karachi. For the development of an eventual scoring system, hospital records of patients admitted to the general surgery service were retrospectively reviewed. These comprised of patients aged 15 years and older who had undergone appendectomy in the time period between October 1995 to April 1998. In total 144 records were looked at and after exclusion of patients with Diabetes Mellitus, malignancy, immunosuppression, lower abdominal pathology/surgery and of records with more than 10% missing data, 401 complete records were finally obtained. Clinical data of these patients was collected using a pre-tested questionnaire, which extracted information on demographics, clinical signs and symptoms and laboratory and radiological investigations. This approach was used to identify parameters that distinguished patients who had a negative appendectomy from those with appendicitis. The potential predictive factors looked at from the patients' records are listed in Table 1.

All analysis was done using Epi Info 6 statistical package. The significance of each of these factors was calculated using chi-square analysis. Those factors with pvalue <0.05 were taken as significant and used for making the scoring system. These were: sex (male), location of initial pain (epigastric), migration of pain to the right lower quadrant, anorexia, vomiting, fever, guarding, rebound tenderness, leukocytosis and neutrophilia. Using Bayesian probability the negative and positive weightage for each

factor was calculated using the following formulae¹¹:

Positive weight = $10 \times \ln(\text{specificity} / \text{sensitivity})$
Negative weight = $10 \times \ln(\text{sensitivity} / \text{specificity})$
When a factor was present a positive weight was given and when it was absent a negative weight was assigned. The weights were rounded off to the nearest integer, applied to the 401 files and summated in order to get the range of most negative score (for positive appendectomy) and most positive score (for negative appendectomy). This turned out to be -83 to +8. This range was then arbitrarily divided into cutoffs taken at increments of 15 i.e., -83, -68, -53, -38, -23, -8, +8. For each cutoff score the sensitivity and specificity was generated using these values. A score with a high specificity and comparable sensitivity was taken as our final cut-off.

The scoring system was then applied to a second population of patients in order to compare suggested actions (derived from the scoring system) to those actually taken by the surgeons. This second population comprised of patients, 15 years and older who presented to the AKUH emergency room with suspected appendicitis in the time period of May 1998 to May 1999. One hundred and twenty six records were obtained and after eliminating patients with Diabetes Mellitus, malignancy, immunosuppression, lower abdominal pathology/surgery and also those records with more than 10% data missing, a final number of 99 records was used. The sensitivity, specificity and accuracy for the level of decision were then calculated.

Results

Of the 401 patients studied retrospectively, 270 (67%) were male and 131 (32%) female. The mean age at presentation was 27 years (15 to 75 years). Of these, 351 (87%) had histologically proven acute appendicitis and 50 (13%) had a normal appendix, resulting in a negative appendectomy rate of 13%. When all 19 potential predictive factors were compared, only 10 were found to occur significantly more often in either of these two groups (Table 1).

Table I. Potential Predictive factors.

1.	Age
2.	Sex*
3.	Location of initial pain*
4.	Migration of pain to right lower quadrant*
5.	Anorexia*
6.	Vomiting*
7.	Diarrhoea
8.	Pain on movement
9.	New onset genitourinary symptoms
10.	Previous episodes of similar pain
11.	Fever*b
12.	Right Lower Quadrant tenderness
13.	Guarding (Right Lower Quadrant)*
14.	Rebound tenderness*
15.	Bowel sounds
16.	Rectal exam : Pelvic tenderness
17.	Rovsing's sign
18.	Leukocytosis*c
19.	Neutrophilia*d

a Liquid faeces or more than three motions per day

b Temperature greater than 37.3 degrees Celsius

c White cell count greater than 10,000 cells per cubic millimetre

d Neutrophil count greater than 75%

***** Statistically significant factors (ten in all)

The positive and negative weights attributed to each significant predictor are listed in Table 2.

Table 2. Significant predictive factors.

	Acute Appendicitis		Normal Appendix		Chi-square (χ^2)	p-value	Sensitivity (%)	Specificity (%)	Weight
	n=351	%	n=50	%					
Sex									
Male	248	71	22	44	14.5	<0.001	70	56	+5
Female	103	29	28	56					-7
Initial Pain									
Epigastric	78	22	2	4	9.1	0.002	22	96	+7
Other	273	78	48	96					-21
Migration to right iliac fossa									
Yes	274	78	24	48	20.72	<0.001	78	52	+5
No	77	22	26	52					-9
Anorexia									
Yes	203	58	9	18	24.12	<0.001	58	82	+12
No	148	42	41	82					-7
Vomiting									
Yes	229	65	21	42	12.27	<0.0004	65	58	+4
No	122	35	29	58					-5
Fever									
Yes	157	45	14	28	5.01	<0.025	44	72	+5
No	194	55	36	72					-3
Guarding									
Yes	236	67	15	30	25.84	<0.001	67	70	+8
No	115	33	35	70					-8
Rebound tenderness									
Yes	298	85	26	52	30.53	<0.001	84	48	+5
No	53	15	24	48					-12
Leukocytosis (per mm ³)									
Yes	311	89	24	48	52.27	<0.001	91	52	+6
No	40	11	26	52					-18
Neutrophilia %									
Yes	275	78	19	38	36.4	<0.001	78	62	+7
No	76	22	31	62					-10

The highest positive predictor was “anorexia” and the highest negative predictor was “initial pain in the epigastric region”.

The diagnostic score in the whole group had a range -100 to +64. However, the range in patients with proven appendicitis was -83 to +64 and in those with a non-inflamed appendix -100 to +8.

Different cut off levels were analyzed for determining an appropriate level for decision-making (Table 3),

Table 3. Potential levels of decision of the scoring system.

	A	B	C	D	E	F	G
Score	-83	-68	-53	-38	-23	-8	8
Sensitivity	1.00	0.99	0.98	0.96	0.86	0.71	0.47
Specificity	0.02	0.22	0.38	0.58	0.86	0.96	1.00

The Score [obtained by the addition of appropriate 'weights' (from Table 2)] is interpreted as follows :

< -83 = Discharge

-83 to -8 = Observe

> -8 = Immediate Surgery

that ranged from a point with maximal sensitivity (Point A) to one with maximal' specificity (Point G) (Figure). For purposes of our analysis, point F was used as a cut-off (Figure).

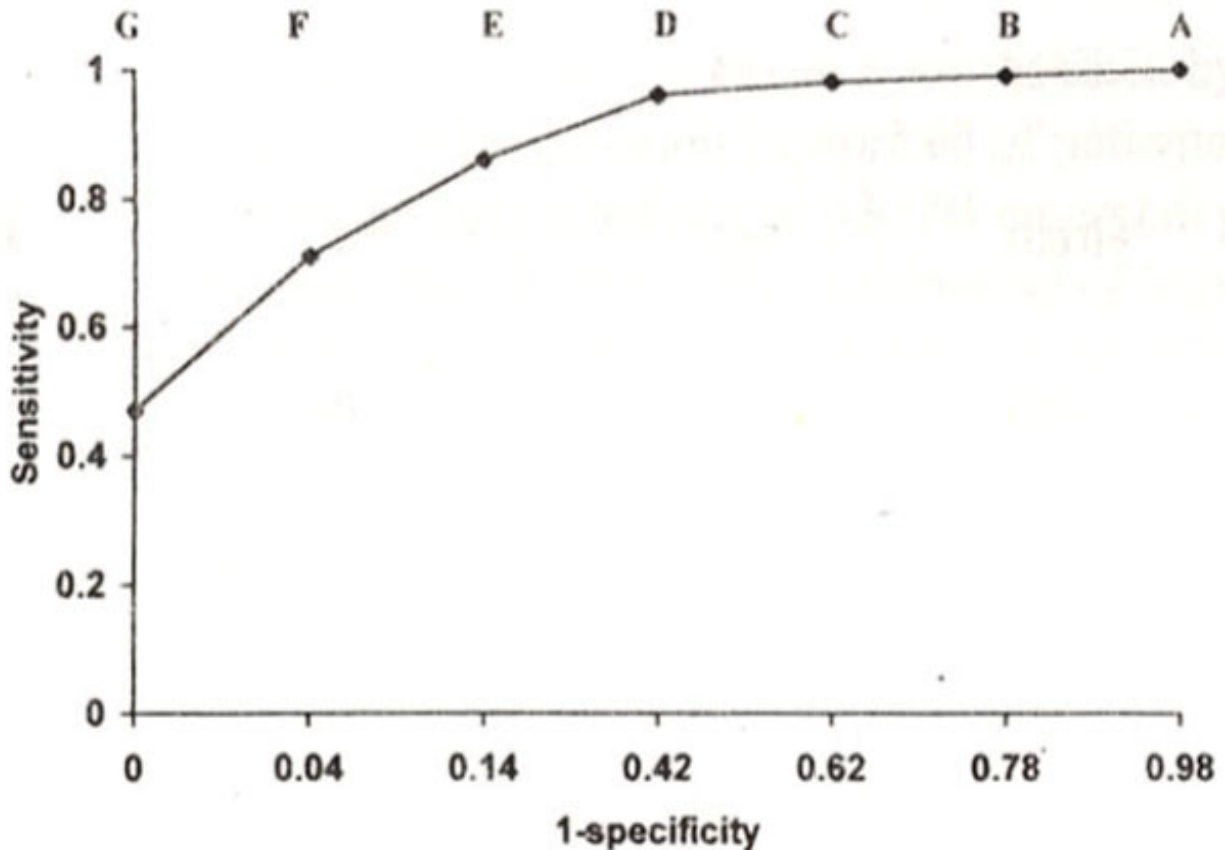


Figure. ROC (Receiver operator characteristic) curve from the data of Table 3. Points A and G represent maximal sensitivity and specificity respectively. Point F was the cut-off selected.

This had a sensitivity of 71% and specificity of 96%. Based on these values, patients with a score greater than -8 were recommended immediate surgery, those with a score less than -83 could be discharged and those with a score between these values could be observed (Table 3).

Of the 99 patients in the second cohort, 99 (93%) were correctly diagnosed by clinicians, with a negative appendectomy rate of 7%.

After applying the scoring system to this second cohort, 65 patients were eligible for immediate surgery, 63 of whom had acute appendicitis (3.1% diagnostic error). There were no patients in the “discharge” group because no one had a score less than -83. Of 33 patients in whom the score suggested active observation, 18 had appendicitis. The accuracy of our scoring system for acute appendicitis was 82%, with a significant difference between men (90%) and women (69%) (Table 4).

Table 4.

Suggested Outcomes	Actual Outcomes			Total
	Positive appendectomy	Negative appendectomy	Observation	
Surgery				
Males	43	1	-	44
Females	20	1	-	21
Observation				
Males	11	5	11	27
Females	7	-	-	7
Discharge				
Males	-	-	-	-
Females	-	-	-	-
Total	81	7	11	99

Sensitivity = 78%

Specificity = 89%

PPV = 97%

NPV = 47%

Thus our scoring system had a sensitivity of 78% and specificity of 89% when applied to this second cohort. The perforation rate from our study was 13%.

Discussion

Appendicitis manifests as a clinical constellation of symptoms. The correct and early diagnosis of appendicitis remains difficult despite the advanced investigations available. In developed countries, the introduction of appendiceal CT appears to have tackled this problem but in developing countries where clinical parameters remain the mainstay of diagnosis, the problem remains³⁻⁷. We believe that the initial assessment of a patient with suspected appendicitis can be improved by the use of a clinical

scoring system. It has been seen that structured preoperative data collection forms can increase the clinical diagnostic accuracy for acute appendicitis¹⁶ as they allow for a more consistent and definitive clinical assessment.

There is growing realization that significant morbidity is associated with negative appendectomies. The 12.5% negative appendectomy rate in our centre is comparably lower than the 20-40% rates from western institutions from the pre-CT era^{14,17-19}.

It should be noted that rates as low as 9% negative appendectomies have been recorded when the paediatric population was excluded¹⁸ as is also the case in our study. Children generally tend to have a higher negative appendectomy rate²⁰. Clinical diagnosis is most reliable in young male patients in whom the rate of negative appendectomy is 10% to 15%. Females of childbearing age have the highest negative appendectomy rates at 35% to 45% because of the clinical overlap between symptoms of appendicitis and gynecological disease. The preoperative application of the score in our study population showed a negative appendectomy rate of 2% in males and 5% in females. These results again reinforce the finding of higher incidence of negative appendectomy in females. However, because the sample size required in calculating this was small these results may not be representative. Previously it was thought that a given rate of negative appendectomy was acceptable so as not to miss a perforated appendix. More recent literature looking at both negative appendectomy and perforation rates Found them to be independent outcomes and not inversely related^{18,21}.

The level of decision of our scoring system for the cut off value of -8 has a sensitivity of 71% and specificity of 96%. A higher specificity was chosen in order to decrease the number of false positives and thus the negative appendectomies. This would lead to a higher false negative rate (22%) and thus put more patients in the observation group rather than sending them directly for surgery. This practice will not adversely affect the patient, as frequent in-hospital re-evaluation will dictate subsequent management. Similar studies done earlier have also chosen a higher specificity for their level of decision for example, 87% in a study done by Ramirez et al¹¹.

When our scoring system was validated on the second cohort, the test sensitivity was calculated at 77% and specificity at 89%. These results are comparable to the Fenyo scoring system which had a sensitivity of 73% and a specificity of 87%²² and superior to the Alvarado scoring system which had a sensitivity of 48% and specificity of 87%²³.

The existing clinical scores appear to have varied results depending on the population on which they are applied^{13,23,24}. Relatively high sensitivity and specificity is recorded when the scoring system is validated on the indigenous population but has poor predictive value when used in other settings²⁵.

Ohmann et al²⁵ applied ten different preexisting clinical scoring systems on a local prospective database and found them to have poor predictive value²⁵. In contrast when Ramirez et al, 1994¹¹ created a new scoring system and tested its accuracy on the same local database, they found a sensitivity of 80% and a specificity of 81%. These results confirm those of other authors²² and suggest that scoring systems developed from a local database can become the ideal complementary method in the diagnosis of suspected acute appendicitis. With a positive predictive value of 97% our scoring system has use as a diagnostic tool for clinicians especially when deciding which patients need further investigations thus leading to better allocation of resources. This applies to patients with equivocal scores in whom, further investigations like ultrasound or CT scan can improve the diagnostic accuracy. Such a scoring system can also help improve data recording if a standardized questionnaire, based on the scoring system, is made part of the initial evaluation. The role of a structured registration form has been emphasized by other authors²¹. The implementation of structured data forms is simple and cost effective. Also in larger surgical units where junior staff with varying clinical experience assess patients with suspected acute appendicitis, the use of such a data form may provide a more systematic approach

to patient management.

Thus in establishing a score based on predictive factors from our own population, we have developed a tool which besides being comparable to existing scoring systems, has shown to significantly reduce the existing negative appendectomy rate from 7 to 2 out of 99. It could, therefore, prove valuable in terms of decreasing unnecessary costs of surgery. This latter aspect merits further research.

The limitations of this study mostly stem from its retrospective methodology. In the first part of the study, when data was collected for devising the score, missing data in files may have biased the final variables. Additionally when incomplete files were eliminated from the final analysis it was assumed that these files were a random selection from the study population. A bias would arise if these eliminated files had incomplete data due to the fact that those patients were more seriously ill and thus there was less time for detailed recording. Further validation of the score may therefore be needed in a prospective manner.

References

1. Khwaja RA, Rasool I, Nadeem IA. Perforated appendicitis versus non-perforated appendicitis. *J. Pak. Med. Assoc.*, 1987;37:325-26.
2. Rehman JS, Auranzeh, Hussain M. Review of acute appendicitis at Civil Hospital Abbottabad *J. Pak. Med. Assoc.*, 1985;35:298-300.
3. Rao PM, Rhea ii, Novelline RA, et al. Effect of computed tomography of the appendix on treatment of patients and use of hospital resources. *NEJM* 1998;338:141-46.
4. Cantrell JR. The diminishing mortality from appendicitis. *Ann. Surg.*, 1995;141 :749-56.
5. Silberman BA. Appendectomy in a large metropolitan hospital:a retrospective analysis of 103 cases. *Am. J. Surg.*, 1981;142:615-61.
6. Temple CL. The natural history of appendicitis in adults: a prospective study. *Ann. Surg.*, 1995;221:278-81.
7. Balthazar EJ, Rofsky NM, Zucker R. Appendicitis: the impact of computed tomography and imaging on negative appendectomy and perforation rates. *AJG*, 1998;93:768-70.
8. Berry J ir, Malt R. Appendicitis near its centenary. *Ann. Surg.*, 1984;200:56775.
9. Velnonch. Balancing the normal appendectomy rate with perforated appendicitis rate: Implication for quality assurance. *Ann. Surg.*, 1992;58:26469.
10. Wets SW, Naylor D. Diagnostic accuracy and short tenn surgical outcomes in cases of suspected acute appendicitis. *J. Can. Med. Asso.*, I 995;152:1617-26.
11. Ramirez JM, Deus J. Practical score to aid decision making in doubtful cases of appendicitis. *Br. J. Surg.*, 1994;81 :680-83.
12. Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann. Emerg. Med.*, 1986;15:557-64.
13. Fenyo O. Routine use of a scoring system for decision making in suspected acute appendicitis in adults. *Acta. Chir. Scand.*, 1987;1 53:545-51.
14. Kalan M, Talbot D, Cunliffe Wi, claP. Evaluation of the modified Alvarado score in the diagnosis of acute appendicitis: a prospective study *Ann. R. Coll. Surg. Eng.*, 1994;76:418-19.
15. Owen TD, William H, Stiff 0, et al. Evaluation of the Alvarado score in acute appendicitis. *J. R. Soc. Med.*, 1992;85:87-88.
16. Komer H, Sondenaa JA, Soreido JA, et al. Structured data collection improves the diagnosis of acute appendicitis. *Br. J. Sug.*, 1998;85:341-44.
17. Calder JDF, Gajraj H. Recent advances in the diagnosis and treatment of acute appendicitis. *Br. J. Hosp. Med.*, 1995;54:129-33.
18. Colson M, Skinner KA, Dunnington 0. High negative appendectomy rates are no longer acceptable.

Am. J. Surg., 1997;174:723-27.

19. Rao PM, Rhea JT, Rattner DW, et al. Introduction of appendiceal CT: impact on negative appendectomy and appendiceal perforation rates. *Ann. Surg.*, 1999;229:344-49.

20. Hale DA, Molloy M, Pearl RH, et al. Appendectomy: a contemporary appraisal. *Ann. Surg.*, 1997;225:252-61.

21. Hale DA, Jacques DP, Molloy M, et al. Appendectomy : improving care through quality improvement. *Arch. Surg.*, 1997;132:153-57.

22. Fenyo O, Lindberg O, Blind P, et al. Diagnostic decision support in suspected appendicitis : validation of a simplified scoring system. *Eur. J. Surg.*, 1997;163:831-38.

23. Gallego MG, Fadrique B, Nieto MA, et al, Evaluation of ultrasonography and clinical diagnostic scoring in suspected appendicitis. *Br. J. Surg.*, 1998;85:3740.

24. Malik AA, Wanie NA. Continuing diagnostic challenge of acute appendicitis. evaluation through modified Alvarado score. *Aust. NZ. J. Surg.*, 1998;68:504-

25. Ohmann C, Yang Q, Franke C. Diagnostic scores for acute appendicitis. Abdominal Pain Study Group. *Eur. J. Surg.*, 1995;161:273-81.