Development of the RIPASA score: a new appendicitis scoring system for the diagnosis of acute appendicitis

Chong C F, Adi M I W, Thien A, Suyoi A, Mackie A J, Tin A S, Tripathi S, Jaman N H, Tan K K, Kok K Y, Mathew V V, Paw O, Chua H B, Yapp S K

Department of Surgery, Raja Isteri Pengiran Anak Saleha Hospital, Bandar Seri Begawan BA1710, Brunei Darussalam

Chong CF, FRCS, FRCSE, MD
Specialist Surgeon

Thien A, MRCS
Senior Medical Officer

Suyoi A, MRCS
Senior Medical Officer

Mackie AJ, FRCS
Senior Medical Officer

Tin AS, MRCS
Senior Medical Officer

Tripathi S, MS
Medical Officer

Jaman NH, MBChB
Medical Officer

Tan KK, FRCS
Specialist Surgeon

Kok KY, FRCS, FAMS
Specialist Surgeon

Mathew VV, FRCS
Specialist Surgeon

Paw O, FRCS
Specialist Surgeon

Chua HB, FRCS
Specialist Surgeon

Yapp SK, FRCS
Specialist Surgeon

Institute of Medicine, University Brunei Darussalam, Jalan Tungku Link, Gadong BE1410, Brunei Darussalam

Adi MIW
Medical Student

Correspondence to:
Mr Chong Chee Fui
Tel: (673) 224 2424 ext 280
Fax: (673) 233 3270
Email: chong_chee_fui@hotmail.com

ABSTRACT

Introduction: Acute appendicitis is one of the most common surgical emergencies. The Alvarado and modified Alvarado scores have been developed to aid diagnosis, but both scoring systems have poor sensitivity and specificity when applied in Middle Eastern and Asian populations. The aim of this study was to develop a new scoring system that is suitable for the local population.

Methods: Clinical data from 312 patients who had undergone an emergency appendicectomy was retrospectively collected and used to generate 15 parameters. The probability was calculated and a score of 0.5, 1.0 or 2.0 was allocated to each parameter. The receiver operating curve (ROC), sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of the new scoring system were derived using the StatsDirect statistical software.

Results: The 15 parameters and the scores generated were age (less than 40 years is 1 point; greater than 40 years is 0.5 point), gender (male is 1 point; female is 0.5 point), right iliac fossa (RIF) pain (0.5 point), migration of pain to RIF (0.5 point), nausea and vomiting (1 point), anorexia (1 point), duration of symptoms (less than 48 hours is 1 point; more than 48 hours is 0.5 point), RIF tenderness (1 point), guarding (2 points), rebound tenderness (1 point), Rovsing's sign (2 points), fever (1 point), raised white cell count (1 point), negative urinalysis (1 point) and foreign national registration identity card (1 point). The optimal cut-off threshold score from the ROC was 7.5, with a sensitivity of 88 percent, a specificity of 67 percent, a PPV of 93 percent and an NPV of 53 percent. The negative appendicectomy rate decreased significantly from 16.3 percent to 6.9 percent, which was a 9.4 percent reduction (p is 0.0007).

Conclusion: The new appendicitis scoring system looked promising when applied to our settings, and had a better sensitivity and specificity than the Alvarado score when applied to Asian populations. A significant reduction in the negative appendicectomy rate was also predicted. A prospective evaluation of this new appendicitis scoring system, referred to as the RIPASA score, is ongoing.

Keywords: acute appendicitis, appendicectomy, diagnostic techniques, surgical, symptoms

Singapore Med J 2010;51(3):220-225

INTRODUCTION

Acute appendicitis is one of the most common surgical emergencies, with a lifetime prevalence rate of approximately one in seven.[1] The incidence is 1.5–1.9 per 1,000 in the male and female population, and is approximately 1.4 times greater in men than in women.[2]

The diagnosis of acute appendicitis is based purely on clinical history and examination combined with laboratory investigations such as elevated white cell count. Despite being a common problem, acute appendicitis remains a difficult diagnosis to establish, particularly among the young, the elderly and females of reproductive age, where a host of other genitourinary and gynaecological inflammatory conditions can present with signs and symptoms that are similar to those of acute appendicitis.[3] A delay in performing an appendicectomy in order to improve its diagnostic accuracy increases the risk of appendicular perforation and sepsis, which in turn increases morbidity and mortality.[4] The opposite is also true, where with reduced diagnostic accuracy, the negative or unnecessary appendicectomy rate is increased, and this is generally reported to be approximately 20%–40%.[5]

Diagnostic accuracy can be further improved through the use of ultrasonography or computed tomography imaging.[6] However, these modalities are costly and may

Table I. Patient demographics.

Demographic	No. (%) (n = 312)
Male: female ratio	180:132
Mean age ± SD (years)	26 ± 13.5
Positive histology for acute appendicitis	261 (83.7)
Negative histology for acute appendicitis	51 (16.3)
Negative appendicectomy rate (%)	16.3
Laparoscopic appendicectomy	42 (13.5)
Mean hospital stay ± SD (days)	4.6 ± 3.8
Postoperative complications (%)	22 (7.1)
Superficial wound infection	13 (4.2)
Wound haematoma	2 (0.6)
Wound pain	4 (1.3)
Intra-abdominal sepsis	3 (1.0)
No. of patients discharged alive	312 (100)

SD: standard deviation

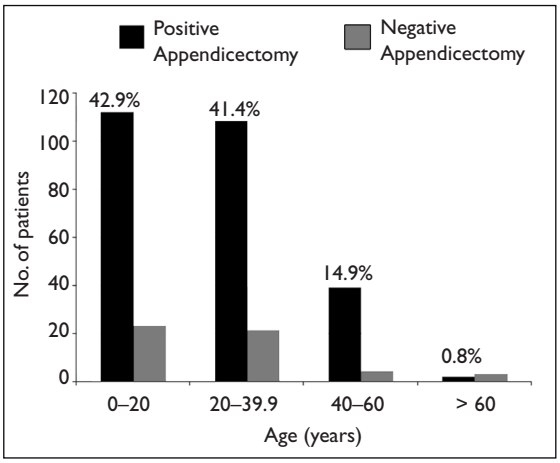

Fig. 1 Distribution of patients who underwent an emergency appendicectomy according to age.

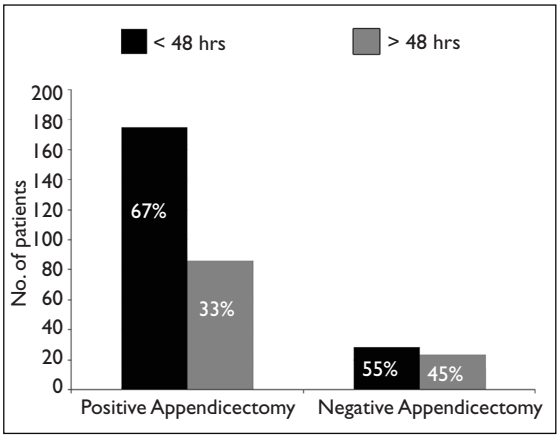

Fig. 2 Distribution of patients who underwent an emergency appendicectomy according to the duration of symptoms.

not be easily available when they are required. Making arrangements for these diagnostic modalities may lead to further delays in diagnosis and surgery. Several scoring systems have been developed to aid in the diagnosis of acute appendicitis. The Alvarado score and the modified Alvarado score are the two most commonly used scoring systems.[5,7] The reported sensitivity and specificity for the Alvarado and the modified Alvarado scores range from 53%–88% and 75%–80%, respectively.[5,7] However, these scoring systems were developed in western countries, and several studies have reported very low sensitivity and specificity when these scores are applied to a population with a completely different ethnic origin and diet.[8,9] Thus, the objective of this study was to develop an appendicitis scoring system that is more applicable to the Southeast Asian region.

METHODS

This was a retrospective study consisting of 400 patients who had undergone an appendicectomy between October 2006 and May 2008, and who were identified from the operation note database of the Department of Surgery, Raja Isteri Pengiran Anak Saleha (RIPAS) Hospital, Brunei Darussalem. The inclusion criteria were patients of all age groups who presented with right iliac fossa (RIF) pain suspected to be acute appendicitis, and who had undergone emergency appendicectomy as the primary procedure. Patients presenting with any form of non-RIF pain, such as lower abdominal pain or right upper quadrant pain, and those who had undergone other emergency laparotomy where appendicectomy was also performed as part of the procedure, or elective appendicectomy, were excluded. Ethical approval for the study was obtained from the Ethics Committee Review Board of RIPAS Hospital.

Of the 400 patients, only 323 had complete medical records, which were collected from the Medical Record Department of RIPAS Hospital. The medical records of the other 77 patients were not traceable, and hence, these patients were excluded from the study. Out of the 323 patients, only 312 patients satisfied our inclusion and exclusion criteria. The other 11 patients were excluded as they presented with non-RIF pain.

The data collected included the patients' demographics (national registration identity card [NRIC] number, age and gender), the presenting symptoms (RIF pain, the migration of pain to the RIF, nausea and vomiting, anorexia and the duration of symptoms), clinical signs (RIF tenderness, guarding, rebound tenderness, Rovsing's sign and fever) and laboratory investigations (elevated white cell count and negative urinalysis). The inclusion of these 15 parameters was agreed upon by a panel of general surgeons at RIPAS Hospital. These 15 parameters form the basis of the new appendicitis scoring system. The probability of each parameter was calculated and scores of 0.5, 1.0 or 2.0

Table II. The probability of acute appendicitis for each parameter, with the scoring of parameters based on probabilities and extra weightage.

Scoring Elements	Probability	Odds ratio	Score	Missing data (%)
Male	0.90	3.10	1.0	0.0
Female	0.75	-	0.5	-
Age < 39.9 yrs	0.83	0.85	1.0	0.0
Age > 40 yrs	0.85	-	0.5	-
RIF pain	0.70	-	0.5	0.0
Migration of RLQ pain	0.83	1.03	0.5	18.0
Anorexia	0.90	0.50	1.0	54.0
Nausea & Vomiting	0.90	0.29	1.0	1.0
Duration of symptoms < 48 hrs	0.86	0.60	1.0	0.0
Duration of symptoms > 48 hrs	0.79	-	0.5	0.0
RIF tenderness	0.84	1.18	1.0	0.3
RIF guarding†	0.92	0.21	2.0	7.0
Rebound tenderness	0.88	0.59	1.0	36.0
Rovsing's Sign†	0.91	0.47	2.0	84.0
Fever	0.94	0.22	1.0	2.0
Raised WCC	0.86	0.42	1.0	0.0
Negative urinalysis*	0.87	0.54	1.0	13.0
Foreign NRIC**	0.96	5.75	1.0	0.0
Minimum Total Score	-	-	2	-
Maximum Total Score	-	-	16	-

† Extra weightage provided by agreement of a panel of general surgeons.
* Negative urinalysis: absence of blood, neutrophils or bacteria.
** Additional parameter.
RIF: right iliac fossa; RLQ: right lower quadrant; WCC: white cell count; NRIC: national registration identity card

points were allocated to each parameter based on its probability, with extra weightage provided to two clinical signs: guarding and Rovsing's signs. Confirmation of acute appendicitis as the final diagnosis was obtained from a histological analysis of the resected appendix at the Department of Histopathology at RIPAS Hospital.

The binomial data was analysed using a non-parametric chi-square test. The probability and odds ratio for each parameter were derived using logistic regression analysis. The receiver operating curve (ROC) at the optimal cut-off threshold score for the new appendicitis scoring system was derived using the StatsDirect statistical software version 2.7.2 (StatsDirect Ltd, Cheshire, UK). The sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) at the optimal cut-off threshold score were also derived from the ROC.[10] The predicted negative appendicectomy rates for the new appendicitis scoring system were also derived and compared with the negative appendicectomy rate from the raw data. The intra-observer and inter-observer variability of the dataset collected were assessed using correlation and regression analysis as well as Bland-Altman plots in ten randomly selected patients.[11]

RESULTS

The study population consisted of 312 patients who had undergone emergency appendicectomy, as shown in Table I. The mean age of the group was 26.0 ± 13.5 years, with a male to female ratio of 180:132 (1.4:1). A positive diagnosis of acute appendicitis was confirmed on histological analysis of the resected appendix in 261 patients, while 51 patients had a normal appendix, indicating a negative appendicectomy rate of 16.3%. The mean duration of hospital stay was 4.6 ± 3.8 days. The rate of postoperative complications was 7%, and consisted mainly of superficial wound infections, as shown in Table I. All 312 patients were discharged alive.

84.3% of the patients with acute appendicitis were < 40 years of age, while 15.7% were > 40 years of age (Fig. 1). Hence, for the development of the new appendicitis scoring system, age was divided into two groups: < 40 years and > 40 years of age. Similarly, the majority of patients with acute appendicitis presented within 48 hours of appearance of symptoms (Fig. 2), and the duration of symptoms in the new appendicitis scoring system was divided into two groups: < 48 hours and > 48 hours.

The parameters included in the new appendicitis scoring system consisted of age, gender, RIF pain, the migration of pain to the RIF, nausea and vomiting, anorexia, the duration of symptoms, RIF tenderness, guarding, rebound tenderness, Rovsing's sign, fever,

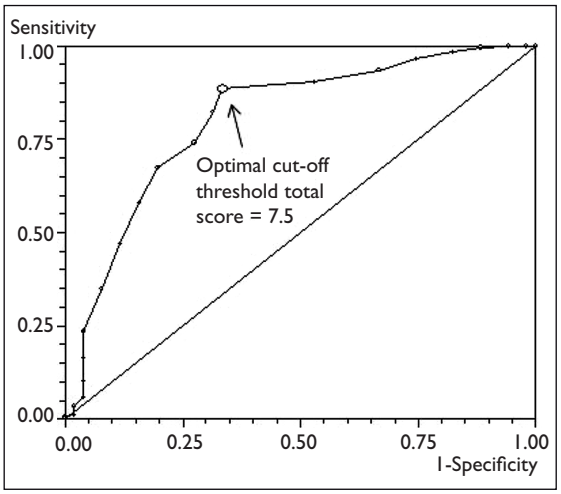

Fig. 3 ROC plot for the new appendicitis scoring system. The optimal cut-off threshold score is 7.5, with a sensitivity and specificity of 0.88 and 0.67, respectively, and a diagnostic accuracy of 0.81. The positive predictive value and negative predictive value are 0.93 and 0.53, respectively.

elevated white cell count, negative urinalysis and a foreign NRIC included as an additional parameter because of the high probability of acute appendicitis seen in foreign nationals presenting with RIF pain (Table II). The probabilities for acute appendicitis were calculated for each of the 15 parameters, as shown in Table II.

Scoring of the parameters was done based on the probability of acute appendicitis. Male gender was found to have a higher probability than female gender, and hence was scored with 1.0 point while female gender was given a score of 0.5 point. As more than 84% of patients with acute appendicitis were < 40 years of age (Fig. 1), despite the slightly lower probability compared with an age > 40 years, having an age < 40 years was scored with 1.0 point, while an age > 40 years was scored with 0.5 point. Both the presence of RIF pain and the migration of pain to RIF were combined for a score of 1.0 point; thus, a score of 0.5 point was allocated to each of these parameters. A duration of symptoms of < 48 hours showed a higher probability of acute appendicitis, and was scored with 1.0 point, while a duration of symptoms > 48 hours was scored with 0.5 point. Both signs of localised guarding and Rovsing's sign were weighted highly by our panel of local general surgeons, as the presence of these two clinical signs was highly indicative of acute appendicitis. Hence, these two parameters were scored with 2.0 points each. The remaining parameters (nausea and vomiting, anorexia, RIF tenderness, rebound tenderness, fever, elevated white cell count, negative urinalysis and foreign NRIC) were all scored with 1.0 point each (see Table II).

The optimal cut-off threshold score derived from the ROC analysis was 7.5, as shown in Fig. 3. Based on this optimal cut-off threshold, the calculated sensitivity and specificity were 88.46% (95% confidence interval [CI] 83.94%–92.08%) and 66.67% (95% CI 52.08%–79.24%), respectively (Fig. 3). The PPV and NPV were 93.00% and 53.00%, respectively (Fig. 3). The diagnostic accuracy was 80.50% (95% CI 73.35%–87.65%) (Fig. 3). The predicted negative appendicectomy rate at the optimal cut-off threshold score of 7.5 was 6.9%, which was a 9.3% reduction from the raw data (16.3%), and this was statistically significant (p = 0.0007).

The correlation regression coefficients for the intra-observer and inter-observer variability analysis of the dataset were 0.93 and 0.88, respectively (Figs. 4a & 5a). Figs. 4b and 5b show the Bland-Altman plots for both intra-observer and inter-observer variability, showing that the majority of the data were within ± 1 standard deviation (SD) of the average difference. Both the correlation regression analysis and Bland-Altman plots indicate that the dataset was reliable.

DISCUSSION

Acute appendicitis is one of the most commonly encountered surgical emergencies, especially by junior doctors on call, with emergency appendicectomy making up 10% of all emergency abdominal surgeries.[12,13] Several scoring systems, such as the Alvarado and modified Alvarado scoring system, have been introduced since 1986 to help with the clinical decision-making process in achieving an accurate diagnosis of acute appendicitis in the fastest and cheapest way.[5,7] However, these two scoring systems were created in the West, and when applied in different environments, such as the Middle East and Asia, the sensitivity and specificity levels achieved were very low.[8,9] Khan et al applied the Alvarado scoring system in an Asian population and only achieved a sensitivity and specificity of 59% and 23%, respectively, with a negative appendicectomy rate of 15.6%.[9] Another study by Al-Hashemy et al in 2004 using the modified Alvarado scoring system in a Middle Eastern population reported a similarly low sensitivity of 53.8% and a specificity of 80%.[8] The sensitivity of the Alvarado score achieved when applied in an oriental population, at the suggested cut-off threshold of 7.0, was similarly low at 50.6%, but achieved a high specificity of 94.5%.[14] However, this improved when the cut-off threshold was lowered to 6.0, with a sensitivity and specificity of 88.3% and 94.5%, respectively, suggesting a definite ethnic difference with regard to the Alvarado score.[14]

Both the Alvarado and modified Alvarado scores

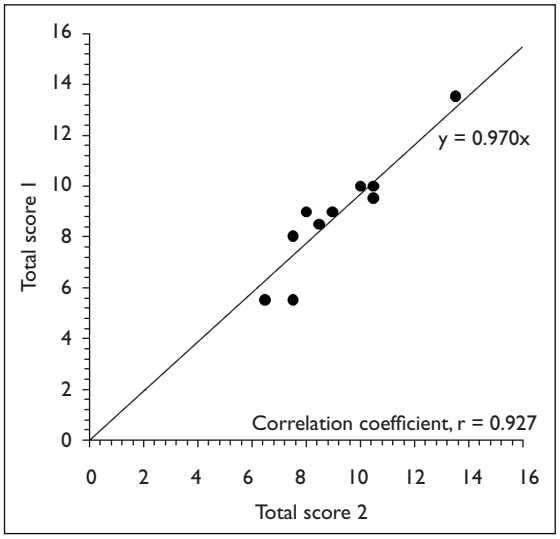

Fig. 4a The correlation regression analysis of intra-observer variability.

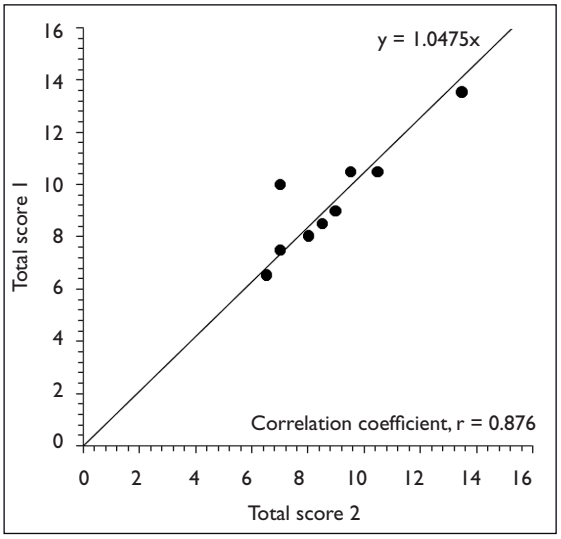

Fig. 5a The correlation regression analysis of inter-observer variability.

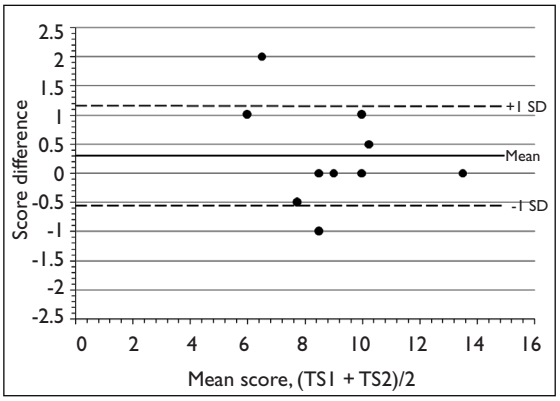

Fig. 4b The Bland-Altman plot of intra-observer variability.

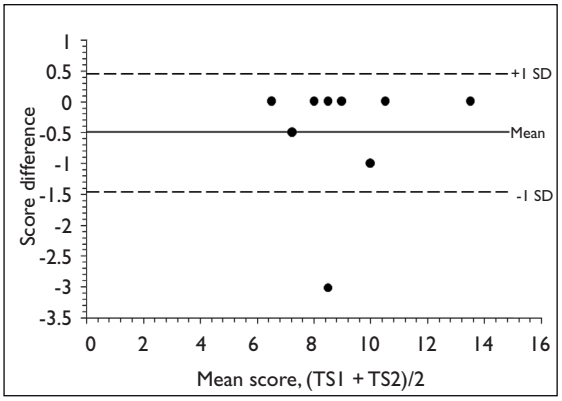

Fig. 5b The Bland-Altman plot of inter-observer variability.

lack parameters that have been shown to be important determinants in the diagnosis of acute appendicitis, such as age, gender and the duration of symptoms. Wani et al have shown that the sensitivity and specificity of the Alvarado scoring system vary with age, gender and the duration of symptoms.[15] Our study has confirmed the presence of age differences (Fig. 1) and differences in the duration of symptoms (Fig. 2) in histologically confirmed cases of acute appendicitis. Furthermore, gender differences in the occurrence of acute appendicitis were also found in our study, with male patients being 1.4 times more likely than female patients to be diagnosed, and this is in keeping with published data.[2] This new appendicitis scoring system includes the three parameters mentioned above as well as four other new parameters deemed important in our local settings, including clinical signs of RIF guarding, Rovsing's sign, negative urinalysis and foreign NRIC status.

Guarding and Rovsing's sign were included as the panel of general surgeons felt that these two clinical signs are earlier indicators of a local inflammatory process such as acute appendicitis, while rebound tenderness is a much later sign when the peritoneum is involved with peritonism. Negative urinalysis was also included to exclude urinary causes of RIF pain, as 60% of our general surgical admission was urological in nature. Lastly, foreign NRIC was included as an additional parameter as the authors had found a high probability (0.8) of acute appendicitis in foreign nationals presenting with RIF pain. There is a large foreign labour workforce in Brunei Darussalam who must pay for their medical treatment at RIPAS Hospital. For this reason, foreign nationals tend to present much later when the symptoms are more severe.

The minimum and maximum total scores achievable with this new appendicitis scoring system were 2 and 16, respectively. The sensitivity and specificity achieved were 88% and 67%, respectively, with a diagnostic accuracy of 81%, which is comparable to the Alvarado score when the latter was applied in a

Western population.[7] This was a definite improvement from the Alvarado score (sensitivity 50.6%–59.0%, specificity 23.0%–94.5%) and modified Alvarado score (sensitivity 53.8%, specificity 80%) when applied to Middle Eastern, Asian or Oriental populations.[8,9,14] The PPV and NPV for the new appendicitis score, at 93% and 53%, respectively, are also comparable to those achieved with the Alvarado and modified Alvarado scores.[8,9,14] Using the new appendicitis scoring system, the predicted negative appendicectomy rate was 6.9%, which was a 9.4% reduction from the raw data, and highly significant statistically ($p = 0.0007$).

This new appendicitis scoring system was specifically developed for our local patient group, but it is likely to be applicable to the South East Asian region, which has populations of similar ethnic origins and diets. The additional parameter of foreign NRIC can be included in the score in countries where there is a large foreign workforce who has to pay for healthcare treatments. This new appendicitis scoring system is easy and simple to apply as the majority of the parameters can be obtained from a routine history and clinical examination.

This study was a retrospective analysis of 312 patients' medical records, and hence, the problem of missing data set is a limitation. As shown in Table II, the missing data set ranged from 0.3% to 84%. Despite this, the sensitivity and specificity derived for this new appendicitis scoring system, when applied to all 312 patients, were comparable to the currently available scoring system.

The new appendicitis scoring system described in this study and referred to as the RIPAS Appendicitis score, or 'RIPASA' score in short, is promising and has good sensitivity, specificity and diagnostic accuracy. It is simple and easy to use, and has been specifically developed for our local patient group, which is reflective of the South East Asian region in terms of diet and ethnic origin. The prospective evaluation of the RIPASA score is ongoing, and the authors aimed to recruit 100–150 patients prospectively.

REFERENCES

1. Stephens PL, Mazzucco JJ. Comparison of ultrasound and the Alvarado score for the diagnosis of acute appendicitis. Conn Med 1999; 63:137-40.
2. Cuscheri A. The small intestine and vermiform appendix. In: Cuschieri A, Giles GR, Mossa AR, eds. Essential Surgical Practice. 3rd ed. Oxford: Butterworth-Heinermann, 1995: 1297-329.
3. Gilmore OJ, Browett JP, Griffin PH, et al. Appendicitis and mimicking conditions. A prospective study. Lancet 1975; 2:421-4.
4. Velanovich V, Satava R. Balancing the normal appendectomy rate with the perforated appendicitis rate: implications for quality assurance. Am Surg 1992; 58:264-9.
5. Kalan M, Talbot D, Cunliffe WJ, Rich AJ. Evaluation of the modified Alvarado score in the diagnosis of acute appendicitis: a prospective study. Ann R Coll Surg Engl 1994; 76:418-9.
6. Baidya N, Rodrigues G, Rao a, Khan SA. Evaluation of Alvarado score in acute appendicitis: a prospective study. Internet J Surg [serial online] 2007:9(1). Available at: www.ispub.com/journal/the_internet_journal_of_surgery/archive/volume_9_number_1.html. Accessed May 1, 2008.
7. Alvarado A. A practical score for the early diagnosis of acute appendicitis. Ann Emerg Med 1986; 15:557-64.
8. Al-Hashemy AM, Seleem MI. Appraisal of the modified Alvarado Score for acute appendicits in adults. Saudi Med J 2004; 25:1229-31.
9. Khan I, ur Rehman A. Application of alvarado scoring system in diagnosis of acute appendicitis. J Ayub Med Coll Abbottabad 2005; 17:41-4.
10. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 1993; 39:561-77.
11. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. Statistician 1983; 32:307-17.
12. Kumar V, Cotran RS, Robbins SL. Appendix. In: Robbin's Basic Pathology. 5th ed. London:WB Saunders, 1992: 520.
13. Pal KM, Khan A. Appendicitis, a continuing challenge. J Pak Med Assoc 1998; 48:189-92.
14. Jang SO, Kim BS, Moon DJ. [Application of alvarado score in patients with suspected appendicitis.] Korean J Gastroenterol 2008; 52:27-31. Korean.
15. Wani MM, Yousaf MN, Khan MA, et al. Usefulness of the Alvarado scoring system with respect to age, sex and time of presentation, with regression analysis of individual parameters. Internet J Surg [serial online] 2007:11(2). Available at: www.ispub.com/journal/the_internet_journal_of_surgery/archive/volume_11_number_2.html. Accessed July 1, 2008.