



A New Approach to Accurate Diagnosis of Acute Appendicitis

Nikolaos E. Tzanakis, M.D.,¹ Stamatis P. Efstathiou, M.D.,² Kecaris Danulidis, M.D.,¹ Georgios E. Rallis, M.D.,¹ Dimitrios I. Tsioulos, M.D.,² Anthimos Chatzivasiliou, M.D.,³ Georgios Peros, M.D.,¹ Nikolaos I. Nikiteas, M.D.⁴

¹Fourth Surgical Clinic, Athens University, Medical School, General State Hospital of Nikea, 3 D.Mantouvalou Str., Piraeus 18454, Greece

²Third Department of Medicine, Athens University, Medical School, Sotiria General Hospital, 152 Mesogion Ave., Athens 11527, Greece

³Department of Radiology, General State Hospital of Nikea, 3 D.Mantouvalou Str., Piraeus 18454, Greece

⁴Second Propeadeutic Department of Surgery, Athens University, Medical School, Laikon Hospital, 17 Ag. Thoma Str, Athens 11364, Greece

⁵Stamatis P. Efstathiou, 19 Fidiou Street, 15562, Athens, Greece

Published Online: August 11, 2005

Abstract. This study aimed (1) to develop a simple scoring system incorporating ultrasound (US) examination and clinical or laboratory predictors for increasing diagnostic accuracy in acute appendicitis (AA), and (2) to evaluate the performance of the scoring system as compared to that of previous models. Fifteen variables including US assessment for patients admitted with suspected AA were considered in multivariate analysis using the finding of AA at operation as the end point (internal study). The new score, together with 11 previous ones, was applied to a prospective independent population of subjects with suspected AA, and the respective performances were compared (external validation study). Among 303 patients (170 males, mean age 28.3 ± 13.3 years) of the internal study, 161 went on to surgery, and 130 had AA at operation. Four independent correlates of AA were identified and used for the derivation of the following integer-based scoring system: number of points = 6 for US demonstrating AA + 4 for tenderness in the right lower quadrant + 3 for rebound tenderness + 2 for leukocyte count $>12,000/\mu\text{l}$. In the external study (201 subjects, 105 males, mean age 28.7 ± 11.9 years, 109 operated, 87 with AA), when the cut-off of ≥ 8 points for AA was used, sensitivity, specificity, accuracy, and area under the curve of the proposed score were 95.4%, 97.4%, 96.5%, and 93%, respectively, exceeding noticeably the previous models. The proposed scoring system introduces a quantitative combination of the clinical evaluation with US imaging and a marker of inflammatory response, which may enhance the diagnostic accuracy for subjects with suspected AA especially in geographical areas where CT scanning is not readily available on a 24-hour basis.

Acute appendicitis (AA) is a common surgical condition of the abdomen, the prompt diagnosis of which is rewarded by a marked decrease in morbidity and mortality [1]. Although the decision to explore a patient with suspected AA is based mainly on disease history and physical findings, the clinical presentation is seldom typical. Therefore diagnostic errors are common, resulting in a median incidence of perforation of 20% and a negative laparotomy rate ranging from 2% to 30% [1].

Correspondence to: Stamatis P. Efstathiou, M.D., e-mail: sefstath@tellas.gr

During the last 20 years, there has been a growing trend toward the use of formal probabilistic reasoning or quantitative data as a guide to clinical decision making. In this respect, several scoring systems, computer-based models, and algorithms [2–12] have been developed for supporting the diagnosis of AA on the basis of grading medical history, clinical symptoms and signs, and indicators of inflammatory response. According to initial evaluation reports, these decision tools are cost-effective and may provide considerable diagnostic aids to physicians [13]. Nevertheless, the aforementioned models have not been routinely applied in general practice because they have failed to achieve adequate accuracy in validation studies [14–17].

Accumulating evidence has suggested that US in experienced hands improves diagnostic accuracy in cases of suspected AA [18, 19]. Thus, sonographic imaging has been proposed as a diagnostic tool even in patients with a clinically high probability of AA, because it accurately depicts a high percentage of normal appendices and alternative diagnoses [20]. However, these findings do not imply that surgeons may not apply their clinical acumen to the management of subjects with suspected AA, inasmuch as series with false-negative sonographic rates of up to 24% have been reported [21]. Furthermore, only scant data exist on the potential combination of US findings with clinical and laboratory variables as an integrated decision tool [22].

The aims of the present study were (1) to develop a simple and reliable scoring system that would incorporate US assessment and particular elements of clinical evaluation and laboratory investigation to provide high diagnostic accuracy in patients with suspected AA and (2) to evaluate the performance of the derived classification rule as compared to that of previously proposed models in a independent database of subjects with suspected AA.

Patients and Methods

The present investigation included overall 504 subjects with suspected AA who were studied in two distinct phases: (1) an internal development study (303 patients) and (2) an external

validation study (201 subjects). Both studies were observational and no intervention was done except for the addition of formalized data collection.

Internal Development Study

The internal study was a prospective analysis conducted between July 1998 and July 2001. Within 1 hour of the clinical and laboratory assessment, all patients admitted with clinically suspected AA underwent sonographic examination of the appendix and the abdomen by a staff radiologist who was blinded to the results of the physical examination and the blood tests, but not to the patient's symptoms. Ultrasound was performed with commercially available high-quality equipment (HDI 3000 unit; Advanced Technology Laboratories, Walpole, MA) with 2–5 MHz curved-array and 5–10 MHz linear-array transducers. Well-established ultrasonographic criteria were applied to discriminate an acutely inflamed appendix from a normal one [19]. All female patients had pelvic examinations. The diagnosis of AA was made exclusively on histopathological grounds by the local pathologist according to previously described standardized criteria [16].

External Validation Study

From August 2001 to August 2003 the score derived from the internal study was applied to an independent database including the next consecutive patients hospitalized for suspected AA. Subsequently, the performance of the score in the above database was compared to that of 11 previously proposed diagnostic scores for AA, which were also calculated by using data from the population of the external study [2–12]. The selection criteria regarding the aforementioned diagnostic scores for AA were (1) development of each score from patients presenting with acute abdominal pain, (2) previous validation in at least one prospective study [15], and (3) feasibility of each score calculation (namely no missing variables) on the basis of the data prospectively collected in our external validation study by using a structured form that included a standardized questionnaire. Because the goal of the present study was to compare the new model with the numerous previous ones, application of the new score to the external study in order to reduce the negative appendectomy rate was not possible without biasing the results. Hence, no score-based intervention took place, and the decision to operate or not was left to the judgment of the senior surgeon, who was not aware of the conclusion of each model for every individual subject. The diagnosis of “non-appendicitis condition” in non-operated participants of both the internal and the external study was supported by telephone surveillance for at least 1 year (median 27 [12–60] months).

Statistical Analysis

Statistical analysis was performed using the Statistical Package for the Social Sciences software (SPSS Inc, Chicago, IL, release 10.0). Acute appendicitis at operation was used as the end point in the internal study. Univariate correlations between the presence of the aforementioned end point and clinical or laboratory features were evaluated with the chi-squared test or Fischer's exact test, as appropriate for categorical data, and with Student's *t*-test or one-way analysis of variance (ANOVA) for continuous variables. A forward stepwise selection procedure, with entry and removal

criteria of $p = 0.05$ and $p = 0.10$, respectively, was used to identify independent predictors of AA. Ninety-five percent confidence intervals (95% CIs) were calculated for each comparison. All tests of significance were two-tailed, and a p value less than 0.05 was considered to be significant. In the external study, the diagnostic performance of each scoring system was tested to define risk groups, reflecting the varying likelihood of AA in an independent population. The Cohen's kappa statistic was calculated for assessing the agreement between scores. Areas under the receiver-operating characteristic (ROC) curves were used to describe the diagnostic performance of each score. The Hosmer-Lemeshow test was used to assess the fit of the models.

Results

Internal Development Study

Among 318 patients admitted with suspected appendicitis, complete data were available for 303 subjects (170 males, 56.1%) who were finally included in the analysis. Participants' mean age was 28.3 ± 13.3 years, 161 patients (52.1%) went on to surgery and 130 (42.9%) had AA at operation. Non-operated subjects were assumed not to have AA, because none of them developed appendicitis during follow-up. Table 1 shows patient characteristics as well as univariate correlates of AA at operation. The final diagnosis for all operated and non-operated subjects in the internal study is presented in Table 2. The negative appendectomy rate was 19.2% (31 out of 161 operated patients). Four independent predictors of the presence of AA were identified in the logistic regression analysis (Table 3). The coefficients (parameter estimates) of the above four factors multiplied by 2 and rounded to the nearest integer, allowed for a simpler re-expression of the final regression model as an integer-based scoring system, which assigned a weight (point) to each predictor and summed the weights of the predictors that were present for a subject: [number of points = 6 for US positive for AA + 4 for tenderness in right lower quadrant + 3 for rebound tenderness + 2 for leukocyte count $> 12,000/\mu\text{l}$]. None of the 22 patients (7.3% of total) who were in the subgroup with the lowest score (0–4 points) had AA, whereas in 126 (96.2%) of the patients with the highest score (8–15 points; $n = 131$ [43.2% of total]), AA was the final diagnosis. Nevertheless, the proportion of subjects with AA among patients with moderate scores (5–7 points; $n = 150$ [49.5% of total]) was very small (4 out of 150, 2.7%). Thus, using the cut-off of ≥ 8 points for the diagnosis of AA in the internal study, a very high probability of AA would have been assigned to subjects with 8–15 points (96.2%, 126/131) as opposed to the very low probability for patients with 0–7 points (2.3%, 4/172).

External Validation Study

The above diagnostic score was calculated for the next 201 patients (105 [52.2%] males, mean age 28.7 ± 11.9 years [range; 15–79 years]) hospitalized for suspected AA. Among the above subjects, 109 (54.2%) went on to surgery and 87 (43.3%) had AA at operation. No significant difference was observed between the populations of the internal study and the external study in the overall frequency of the four above-mentioned independent predictors, as well as in terms of the remaining clinical and demographic characteristics. The application of the new classification tool to the external database showed 96.5% of subjects with

Table 1. Demographic, clinical, and laboratory characteristics of 303 subjects with suspected appendicitis in the internal study and univariate correlates of acute appendicitis at operation (OR, odds ratio; 95% CI, 95% confidence intervals).

Patients' characteristics (<i>n</i> = 303)	Acute appendicitis (%) [<i>n</i> = 130, 42.9%]	No appendicitis (%) [<i>n</i> = 173, 57.1%]	<i>p</i> value	OR	95% CI
<i>Demographic data</i>					
Male sex	78 (59.5)	92 (53.5)	0.292	1.3	0.8, 2.1
Age [mean ± SD (range)]	27.2 ± 12.2 (15–85)	29.4 ± 14.7 (15–86)	0.889	2.2 ± 2.1 ^a	–4.9, 4.7
<i>Symptoms</i>					
Anorexia	90 (68.7)	122 (70.9)	0.675	1.1	0.6, 1.8
Vomiting	62 (47.3)	64 (37.2)	0.076	1.6	0.9, 2.4
Migration of pain	79 (60.3)	57 (33.1)	<0.001	3.1	1.9, 4.9
Duration of symptoms < 48 hours	109 (83.2)	112 (65.1)	<0.001	2.7	1.5, 4.6
<i>Signs</i>					
Tenderness in right lower quadrant	117 (89.3)	71 (41.3)	<0.001	11.9	6.3, 22.4
Rebound tenderness	86 (65.6)	44 (25.6)	<0.001	5.6	3.4, 9.1
Guarding	64 (48.8)	33 (19.2)	<0.001	4.1	2.4, 6.7
Rectal tenderness	35 (26.7)	55 (31.9)	0.321	1.3	0.8, 2.1
<i>Laboratory data</i>					
Leukocyte count > 12,000/μl	79 (60.3)	18 (10.5)	<0.001	13.1	7.1, 23.7
Neutrophils > 75%	111 (84.7)	78 (45.4)	<0.001	6.7	3.8, 11.7
CRP > 9 mg/dl	105 (80.2)	80 (46.5)	<0.001	4.6	2.8, 7.8
Temperature > 37.5°C	96 (73.3)	101 (58.7)	0.008	2.1	1.2, 3.3
Ultrasound positive for acute appendicitis	107 (81.7)	9 (5.2)	<0.001	80.7	36.1, 180.4

^aMean difference ± SD.

OR: odds ratio; CRP: C-reactive protein

Table 2. Final diagnosis for 303 subjects with suspected appendicitis in the internal study.

Operated patients (<i>n</i> = 161)	No.	Non operated patients (<i>n</i> = 142)	No.
Diagnosis		Diagnosis	
<i>Positive appendectomy</i>			
Simple appendicitis	130 (80.8%)	Nonspecific abdominal pain	107
Suppurative appendicitis	70	Gastroenteritis	11
Abscessed appendicitis	27	Gynecologic conditions	9
Gangrenous appendicitis	2	Liver and gallbladder disease	7
Perforated appendicitis	17	Diverticular disease	3
	14	Urolithiasis	3
<i>Negative appendectomy</i>			
Nonspecific abdominal pain	31 (19.2%)	Pneumonia	2
Acute mesenteric adenitis	16		
Acute gynecological disease	4		
Diverticulitis	4		
Omental torsion	2		
Perforated duodenal ulcer	2		
Urinary tract infection	1		

8–15 points to have AA (Table 4). The proposed diagnostic model yielded a score of < 8 points for all 92 non-operated patients in the external study. The level of agreement of the proposed score as estimated by the kappa statistic was high with Eskelinen and Ohmann scores and moderate to fair with the remaining ones (Table 5). The present model exceeded noticeably the previous ones in diagnostic accuracy (Table 5) as well as in discriminatory capacity as expressed by area under the curve (AUC) ($p < 0.001$; Table 6, Fig. 1).

Discussion

The model developed in the present study combines the diagnostic value of four variables: namely two well-recognized clinical features of AA (tenderness in the right lower quadrant and rebound tenderness) [1], US imaging, and leukocytosis, the latter reflecting the inflammatory response. The prominence of the aforementioned factors as independent correlates of AA corroborates previous reports, which have shown scores not including the above clinical variables and leukocytosis to provide poorer

discrimination [1, 15]. With regard to the varied weighting of the four multivariate predictors, a positive US finding surpassed any other factor by introducing an at least 5.5-fold increase to the probability of AA as suggested by 95% CIs (Table 3).

According to the proposed threshold of ≥ 8 points, if the appendix is sonographically shown to be inflamed, the presence of at least one additional factor is required to establish AA, whereas in the absence of US demonstrating AA, all three remaining variables are necessary for the diagnosis. For example, the above model would suggest the diagnosis of AA in a patient with leukocytosis and a positive US finding (total score 8 points), even if rebound or right lower quadrant tenderness were lacking. The application of the new system to the external database yielded an impressive diagnostic accuracy of 96.5%, which exceeded noticeably the performance of previous scores, whereas the comparison of the corresponding AUCs showed a clearly greater discriminatory capacity of the present score (Table 6, Fig. 1), 95% CIs excluding an AUC for the proposed model smaller than 0.93. The superiority of the new score could be attributed to the incorporation of an imaging modality in a formal decision tool for AA,

Table 3. Multivariate analysis of predictors of acute appendicitis at operation in the internal study (OR, odds ratio; 95% CI, 95% confidence intervals).

Variables	Parameter estimate	Standard error	p value	OR	95% CI
Ultrasonography positive for appendicitis	2.998	0.609	< 0.001	17.1	5.5, 41.2
Tenderness in right lower quadrant	2.012	0.588	< 0.001	8.4	2.4, 26.4
Rebound tenderness	1.549	0.599	< 0.001	5.7	1.8, 19.5
Leukocyte count > 12,000/ μ l	1.121	0.562	0.009	3.6	1.3, 12.2
Migration of pain	0.454	0.961	0.102	1.5	0.6, 10.2
Guarding	0.281	0.473	0.159	1.4	0.5, 7.3
Neutrophils > 75%	0.255	0.441	0.288	1.4	0.5, 4.8
CRP > 9 mg/dl	0.223	0.418	0.359	1.3	0.4, 4.2
Temperature > 37.5°C	0.172	0.235	0.631	1.2	0.4, 3.9
Duration of symptoms < 48 hours	0.139	0.112	0.784	1.2	0.2, 3.5

Table 4. Performance of the proposed diagnostic score in the external validation study.

Number of points	Number of patients (% of total)	Acute appendicitis	Non appendicitis condition	Percentage of patients with appendicitis among patients with the respective score
0-4	10 (5.0%)	0	10	0%
5-7	105 (52.2%)	4	101	3.8%
8-15	86 (42.8%)	83	3	96.5%

Table 5. Comparison of the proposed score with the previous ones.

Scoring instrument	SNS	SPC	PPV	NPV	PLR	NLR	ACR	True (+)	False (+)	True (-)	False (-)	Kappa
<i>Van Way</i>	81.6	74.6	71.0 (61.5-78.9)	84.2 (75.8-90.0)	3.21	0.25	77.6	71	29	85	16	0.56
<i>Teicher</i>	88.5	83.3	80.2 (71.1-86.9)	90.5 (83.4-94.7)	5.31	0.14	85.6	77	19	95	10	0.64
<i>Arnbjörnsson</i>	81.6	71.1	68.3 (58.8-76.4)	83.5 (74.9-89.6)	2.81	0.26	75.6	71	33	81	16	0.49
<i>Alvarado</i>	89.7	76.3	74.3 (65.2-81.7)	90.6 (83.1-94.9)	3.78	0.14	82.1	78	27	87	9	0.58
<i>Fenyö</i>	90.8	85.1	82.3 (73.5-88.6)	92.4 (85.7-96.1)	6.09	0.11	87.6	79	17	97	8	0.70
<i>Lindberg</i>	85.1	87.7	84.1 (75.1-90.3)	88.5 (81.3-93.2)	6.92	0.17	86.6	74	14	100	13	0.66
<i>Izbicki</i>	80.5	70.2	67.3 (57.8-75.6)	82.5 (73.7-88.8)	2.69	0.28	74.6	70	34	80	17	0.47
<i>de Dombal</i>	80.5	72.8	69.3 (59.7-77.5)	83.0 (74.5-89.1)	2.96	0.27	76.1	70	31	83	17	0.51
<i>Christian</i>	85.1	85.1	81.3 (72.1-87.9)	88.2 (80.8-92.9)	5.71	0.18	85.1	74	17	97	13	0.65
<i>Eskelinen</i>	82.8	92.1	88.9 (80.2-94.1)	87.5 (80.4-92.3)	10.48	0.19	88.1	72	9	105	15	0.78
<i>Ohmann</i>	93.1	83.3	81.0 (72.2-87.5)	94.1 (87.6-97.2)	5.58	0.08	87.6	81	19	95	6	0.82
<i>Proposed score</i>	95.4	97.4	96.5 (90.2-98.8)	96.5 (91.4-98.6)	36.3	0.05	96.5	83	3	111	4	—

SNS: sensitivity; SPC: specificity; PPV: positive predictive value; NPV: negative predictive value; PLR: positive likelihood ratio; NLR: negative likelihood ratio; ACR: accuracy.

Kappa statistic indicates the level of agreement of the proposed score with the previous ones.

Values in parentheses represent 95% confidence intervals for predictive values.

which is the novel diagnostic procedure introduced in the present study.

Although sonographic imaging of the abdomen has been established as a useful tool in diagnosis of AA being of particular value in patients with atypical presentation [23], its accuracy has been doubted in more recent large studies and meta-analyses [18, 19, 21, 24-26]. In this respect, it has been demonstrated that, when US is used as the determining factor for operative therapy, it cannot be relied on to the exclusion of the surgeon's careful and repeated evaluation [21]. Furthermore, a prospective multicenter observational trial on 2280 patients with acute abdominal pain reported no correlation between the sonographic findings of the appendix and the diagnostic accuracy of the clinician, the rate of negative appendectomy, and the perforation rates, thus suggesting no clear benefit of US scanning of the appendix in the routine clinical setting [19]. In addition, sonography failed to improve the diagnostic accuracy or the negative appendectomy rate and was even found to delay surgical consultation and appendectomy in a large study that included 766 subjects [24]. Nevertheless, it has

Table 6. Discriminatory power of the proposed score as well as of 11 previous scores for diagnosis of acute appendicitis expressed as areas under the receiver-operating characteristic curves (95% CI, 95% confidence intervals:).

Score	Area under the curve	Standard error	p Value	95% CI
<i>Van Way</i>	0.78*	0.03	< 0.001	0.71, 0.85
<i>Teicher</i>	0.86*	0.03	< 0.001	0.80, 0.92
<i>Arnbjörnsson</i>	0.76*	0.04	< 0.001	0.69, 0.83
<i>Alvarado</i>	0.83*	0.03	< 0.001	0.77, 0.89
<i>Fenyö</i>	0.88*	0.02	< 0.001	0.83, 0.93
<i>Lindberg</i>	0.86*	0.02	< 0.001	0.81, 0.92
<i>Izbicki</i>	0.75*	0.04	< 0.001	0.68, 0.82
<i>de Dombal</i>	0.77*	0.04	< 0.001	0.69, 0.83
<i>Christian</i>	0.85*	0.03	< 0.001	0.79, 0.91
<i>Eskelinen</i>	0.87*	0.03	< 0.001	0.82, 0.93
<i>Ohmann</i>	0.88*	0.02	< 0.001	0.83, 0.94
<i>Proposed score</i>	0.96	0.02	< 0.001	0.93, 0.99

p < 0.001 for differences between the area of each preexisting score and that of the new score.

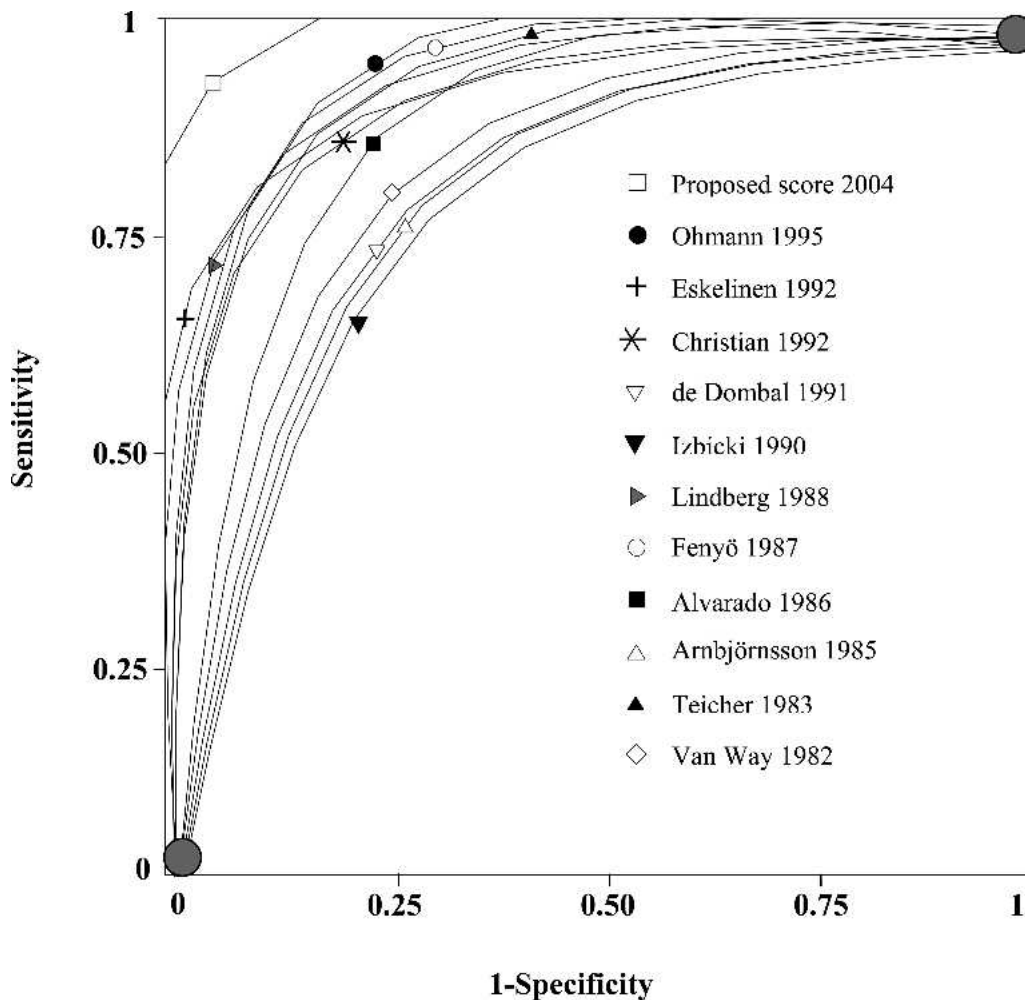


Fig. 1. Receiver-operating characteristic curves plotted on the basis of the proposed score as well as of 11 previous scores for diagnosis of acute appendicitis.

been shown that US is unnecessary when there is a high degree of clinical suspicion as expressed by a positive Alvarado score, whereas the additional information provided by US improves diagnostic accuracy in the case of a negative or equivocal Alvarado score [25]. Moreover, a meta-analysis published in the middle 1990s suggested that US is most helpful in patients with an indeterminate probability of the disease after the initial evaluation and should not be used to exclude AA in subjects with classic signs and symptoms because of the underlying relatively high false-negative rate [18]. Finally, a more recent meta-analysis on the value of US in the diagnosis of AA revealed disappointing results in multi-center trials, suggesting that the adequate performance of sonography in single-center studies may not reflect surgical everyday life [26].

Ultrasound is rapid, noninvasive, inexpensive, and requires no patient preparation or contrast material administration [23]. Because it involves no ionizing radiation and excels in the depiction of acute gynecologic conditions, it is recommended as the initial imaging study in children [27] and in women [28], especially during pregnancy [29]. Yet, the limitations of US include its reduced accuracy in obese or muscular subjects, as well as in patients with perforated AA (approximately 50%) compared to that observed in nonperforated AA (80%) [23]. Furthermore, US is known to be highly operator-dependent, the learning curve required to develop the technique for sonographically scanning the

right lower quadrant is considerable, and there are many interpretive pitfalls to be avoided [23]. It has been shown, however, that even if radiology residents or inexperienced surgeons conduct the imaging, the accuracy of US is not diminished [30, 31]. In any case, although the criteria for the US-based diagnosis of AA are well-established and reliable, the inexperienced examiner, working with poor equipment and/or technique, will provide suboptimal results, and this possibility should be taken into account when incorporating sonographic criteria in the diagnostic pattern.

The use of US in the setting of suspected AA might be questioned in an era when appendiceal computed tomography (CT) has been demonstrated to provide an accuracy rate as high as 98% in the diagnosis of AA, leading to improved patient care and reduced use of hospital resources [32]. Moreover, CT has repeatedly been shown to exhibit superior discriminatory capacity compared to US in both adults and adolescents with suspected AA [33–35], suggesting that the proposed classification system may not apply to geographical areas where CT scanning is readily available on a 24-hour basis. In this study, the inability to routinely perform CT scanning may account to a great extent for the relatively high false positive rate of approximately 20%. This number of false positive diagnoses would be unacceptable in most Westernized nations, where the appropriate CT utilization in community hospitals has been shown to reduce the negative

appendectomy rate from 14%–20% to 2%–7% [36–38]. Nevertheless, because many portions of the world health community may still not be able to afford CT scanning but can afford US equipment, the combined systematic implementation of sonographic evaluation and clinical acumen could be valuable as suggested by the present study.

Because the simultaneous application of the preexisting models and the new score to the same database has favored the latter, the respective clinical implications should be further evaluated. A prospective interventional large-scale evaluation in different clinical environments, in an adequate controlled study comparing a baseline phase without scoring to a subsequent phase with scoring would probably be the optimal approach [15, 16]. To reduce bias with such a design, uniform data collection should be carried out according to constant definitions, with standardized performance criteria used to ensure objective evaluation [16].

Any diagnostic support for AA should be warmly welcomed if it has been proven to be clinically valuable, because unacceptably high negative appendectomy and perforation rates are still reported in many portions of the world health community. However, apart from being familiar with elements not included in a quantitative model, physicians may be able to provide superior imputations of missing data for an individual patient and to integrate the diagnostic estimate as part of their overall patient assessment. Therefore, including the proposed score in the diagnostic procedure is worth trying and may enhance a surgeons discriminatory capacity, under the prerequisite that it will be considered as an adjunct in decision making that cannot supplant careful surgical judgment.

References

- Andersson RE. Meta-analysis of the clinical and laboratory diagnosis of appendicitis. *Br. J. Surg.* 2004;91:28–37
- Van Way CW 3rd, Murphy JR, Dunn EL, et al. A feasibility study of computer aided diagnosis in appendicitis. *Surg. Gynecol. Obstet.* 1982;155:685–688
- Teicher I, Landa B, Cohen M, et al. Scoring system to aid in diagnoses of appendicitis. *Ann. Surg.* 1983;198:753–759
- Arnbjörnsson E. Scoring system for computer-aided diagnosis of acute appendicitis. The value of prospective versus retrospective studies. *Ann. Chir. Gynaecol.* 1985;74:159–166
- Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann. Emerg. Med.* 1986;15:557–564
- Fenyö G. Routine use of a scoring system for decision-making in suspected acute appendicitis in adults. *Acta Chir. Scand.* 1987;153:545–551
- Lindberg G, Feny G. Algorithmic diagnosis of appendicitis using Bayes' theorem and logistic regression. In Bernardo JM, DeGroot MH, Lindley DV, Smith AF, editors. *Bayesian Statistics 3*, Oxford University Press, 1988, pp 665–668
- Izbicki JR, Wilker DK, Mandelkow HK, et al. Retro- and prospective studies on the value of clinical and laboratory chemical data in acute appendicitis. *Chirurg* 1990;61:887–894
- de Dombal FT. *Diagnosis of Acute Abdominal Pain*. Churchill Livingstone: Edinburgh, 1991, pp 105–106
- Christian F, Christian GP. A simple scoring system to reduce the negative appendectomy rate. *Ann. R. Coll. Surg. Engl.* 1992;74:281–285
- Eskelinen M, Ikonen J, Lipponen P. A computer-based diagnostic score to aid in diagnosis of acute appendicitis: a prospective study of 1333 patients with acute abdominal pain. *Theor. Surg.* 1992;7:86–90
- Ohmann C, Franke C, Yang Q, et al. Diagnostic score for acute appendicitis. *Chirurg* 1995;66:135–141
- Hoffmann J, Rasmussen OO. Aids in the diagnosis of acute appendicitis. *Br. J. Surg.* 1989;76:774–779
- Kalan M, Talbot D, Cunliffe WJ, et al. Evaluation of the modified Alvarado score in the diagnosis of acute appendicitis: a prospective study. *Ann. R. Coll. Surg. Engl.* 1994;76:418–419
- Ohmann C, Yang Q, Franke C. Diagnostic scores for acute appendicitis. *Abdominal Pain Study Group. Eur. J. Surg.* 1995;161:273–281
- Ohmann C, Franke C, Yang Q. Clinical benefit of a diagnostic score for appendicitis: results of a prospective interventional study. *German Study Group of Acute Abdominal Pain. Arch. Surg.* 1999;134:993–996
- Sitter H, Hoffmann S, Hassan I, et al. Diagnostic score in appendicitis. Validation of a diagnostic score (Eskelinen score) in patients in whom acute appendicitis is suspected. *Langenbecks. Arch. Surg.* 2004;389:213–218
- Orr RK, Porter D, Hartman D. Ultrasonography to evaluate adults for appendicitis: decision making based on meta-analysis and probabilistic reasoning. *Acad. Emerg. Med.* 1995;2:644–650
- Franke C, Bohner H, Yang Q, et al. Ultrasonography for diagnosis of acute appendicitis: results of a prospective multicenter trial. *Acute Abdominal Pain Study Group. World J. Surg.* 1999;23:141–146
- Rettenbacher T, Hollerweger A, Gritzmann N, et al. Appendicitis: should diagnostic imaging be performed if the clinical presentation is highly suggestive of the disease?. *Gastroenterology* 2002;123:992–998
- Wade DS, Marrow SE, Balsara ZN, et al. Accuracy of ultrasound in the diagnosis of acute appendicitis compared with the surgeon's clinical impression. *Arch. Surg.* 1993;128:1039–1046
- Zielke A, Hasse C, Sitter H, et al. Influence of ultrasound on clinical decision making in acute appendicitis: a prospective study. *Eur. J. Surg.* 1998;164:201–209
- Birnbaum BA, Wilson SR. Appendicitis at the millennium. *Radiology* 2000;215:337–348
- Lee SL, Walsh AJ, Ho HS. Computed tomography and ultrasonography do not improve and may delay the diagnosis and treatment of acute appendicitis. *Arch. Surg.* 2001;136:556–562
- Stephens PL, Mazzucco JJ. Comparison of ultrasound and the Alvarado score for the diagnosis of acute appendicitis. *Conn. Med.* 1999;63:137–140
- Obermaier R, Benz S, Asgharnia M, et al. Value of ultrasound in the diagnosis of acute appendicitis: interesting aspects. *Eur. J. Med. Res.* 2003;8:451–456
- Puig S, Hormann M, Rebhandl W, et al. US as a primary diagnostic tool in relation to negative appendectomy: six years experience. *Radiology* 2003;226:101–104
- Bendeck SE, Nino-Murcia M, Berry GJ, et al. Imaging for suspected appendicitis: negative appendectomy and perforation rates. *Radiology* 2002;225:131–136
- Lim HK, Bae SH, Seo GS. Diagnosis of acute appendicitis in pregnant women: value of sonography. *A.J.R. Am. J. Roentgenol.* 1992;159:539–542
- Chesbrough RM, Burkhard TK, Balsara ZN, et al. Self-localization in US of appendicitis: an addition to graded compression. *Radiology* 1993;187:349–351
- Amgwerd M, Rothlin M, Candinas D, et al. Ultrasound diagnosis of appendicitis by surgeons: a matter of experience? A prospective study. *Langenbecks Arch. Chir.* 1994;379:335–340
- Rao PM, Rhea JT, Novelline RA, et al. Effect of computed tomography of the appendix on treatment of patients and use of hospital resources. *N. Engl. J. Med.* 1998;338:141–146
- Pickuth D, Heywang-Kobrunner SH, Spielmann RP. Suspected acute appendicitis: is ultrasonography or computed tomography the preferred imaging technique? *Eur. J. Surg.* 2000;166:315–319
- Wise SW, Labuski MR, Kasales CJ, et al. Comparative assessment of CT and sonographic techniques for appendiceal imaging. *A.J.R. Am. J. Roentgenol.* 2001;176:933–941
- Terasawa T, Blackmore CC, Bent S, et al. Systematic review: computed tomography and ultrasonography to detect acute appendicitis in adults and adolescents. *Ann. Intern. Med.* 2004;141:537–546
- Schuler JG, Shortsleeve MJ, Goldenson RS, et al. Is there a role for abdominal computed tomographic scans in appendicitis? *Arch. Surg.* 1998;133:373–376
- Rao PM, Rhea JT, Rattner DW, et al. Introduction of appendiceal CT: impact on negative appendectomy and appendiceal perforation rates. *Ann. Surg.* 1999;229:344–349
- Jones K, Pena AA, Dunn EL, et al. Are negative appendectomies still acceptable? *Am. J. Surg.* 2004;188:748–754