Comparison of Clinical Judgment and Diagnostic Ultrasonography in the Diagnosis of Acute Appendicitis: Experience with a Score-aided Diagnosis

Henrik Jahn,[1] Finn K. Mathiesen,[2] Kirsten Neckelmann,[1] Claus P. Hovendal,[1] Torben Bellstrøm[2] and Finn Gottrup[3]

From the Departments of [1]Surgery and [2]Radiology, Odense University Hospital, Odense, Denmark, and [3]Copenhagen Wound Healing Center, Bispebjerg University Hospital, Copenhagen, Denmark

Eur J Surg 1997; 163: 433–443

ABSTRACT

Objective: To evaluate the diagnostic accuracy of clinical judgment and diagnostic ultrasonography (US) used routinely and to create a scoring system to aid diagnosis.

Design: Prospective, double-blind study.

Setting: University hospital, Denmark.

Subjects: 222 Consecutive patients suspected of having acute appendicitis admitted between 0800 and midnight from June 1990 to June 1992.

Interventions: 148 Patients (67%) underwent appendicectomy and the remaining 74 patients were observed. 193 Patients (87%) had a diagnostic US examination. 21 Predictive variables were collected prospectively to create a scoring system.

Main outcome measures: Results of surgical pathological findings, clinical outcome (observed group), diagnostic US, and values of diagnostic score.

Results: The decision to operate was made by a junior surgeon solely on the clinical examination, which yielded a diagnostic accuracy of 76%, specificity of 58%, and negative appendicectomy rate of 36%. 193 Patients underwent diagnostic US conducted by the radiologist on call of whom 123 were operated on, 78 for histologically proven appendicitis. US had a diagnostic accuracy of 72%, sensitivity of 49%, and specificity of 88%. Of the 21 predictive factors for acute appendicitis 11 were significant ($p < 0.05$): total white cell count (WCC) ($>10 \times 10^9/1$), migration of pain to the right lower quadrant, gradual onset of pain, increasing intensity of pain, pain aggravated by movement, pain aggravated by coughing, anorexia, vomiting, indirect tenderness (Rovsing's sign), muscle spasm, and sex. These 11 predictors were assigned an appropriate weight, based on the likelihood ratio, and used to create a scoring system. The score performed poorly if it was used to separate patients for observation and those for appendicectomy. However, if the score was used with two cut-off points resulting in three test zones (low, intermediate, and high risk of having acute appendicitis), some diagnostic benefit was seen for those patients within the zones of high and low probability.

Conclusion: The clinical judgment of a junior surgeon was disappointing, and diagnostic aids are desirable to reduce the negative appendicectomy rate. Diagnostic US performed poorly as a routine procedure. Application of an up to date scoring system might be of some help to patients with a high or low probability of acute appendicitis, but any conclusion about its clinical application cannot be drawn from this study.

Key words: acute appendicitis, appendicectomy, diagnostic accuracy, sonographic diagnosis, ultrasonography, diagnostic score, scoring system, prospective study.

INTRODUCTION

Acute appendicitis is a common and serious abdominal disorder with a potentially lethal outcome. Negative appendicectomy rates between 25% and 40% (3) and perforation rates of nearly 37% (17, 18) have been published. These rates illustrate the diagnostic dilemma of acute appendicitis as well as the various attitudes among surgeons toward the proper time for surgical intervention. A negative appendicectomy rate of 15%–30% has been accepted to be the clinical standard (7). However, in recent years, several reports have been published giving morbidity rates of 5%–15% after a negative exploration, which are not significantly different from non-perforated appendicitis (2, 7, 18). These observations indicated that negative surgical exploration is not harmless, and brought the liberal attitude towards diagnosis and management into disrepute. Many efforts to improve the diagnostic accuracy have been made, and with the present diagnostic methods a negative surgical exploration rate of 15%–30% is no longer acceptable (7).

In an attempt to improve the diagnostic accuracy in

 ISSN 1102–4151

Table I. *Twenty one potential predictors of acute appendicitis in 222 patients with suspected appendicitis*

Figures are percentages of patients, and weight of score.

Variable	Acute appendicitis ($n = 94$)	Normal appendix ($n = 128$)	Weight of score
Rectal temperature (°C):			
≤37.4	30	45	0
37.5-38.4	60	46	0
≥38.5	10	9	0
Pulse (bpm):			
≤89	69	76	0
≥90	31	24	0
White cell count ($\times 10^9$/L):*			
≤10	17	68	−14
>10	83	32	+10
Duration of pain (hours):			
<13	15	19	0
13-48	71	59	0
>48	14	22	0
Pain migrated to right lower quadrant:*			
Yes	72	38	+6
No	28	62	−8
Onset of pain:*			
Sudden	12	25	−7
Gradual	88	75	+2
Intensity of pain:*			
Increasing	84	52	+5
Decreasing or unchanged	16	48	−11
Pain aggravated by movement:*			
Yes	88	68	+3
No	12	32	−10
Pain aggravated by coughing:*			
Yes	85	62	+3
No	15	38	−9
Anorexia:*			
Yes	84	66	+2
No	16	34	−8
Nausea:			
Yes	68	60	0
No	32	40	0
Vomiting:*			
Yes	49	31	+5
No	51	69	−3
Raised temperature before admission (°C):			
≤37.5	55	45	0
>37.5	45	55	0
Shivering:			
Yes	38	35	0
No	62	65	0
Diarrhoea:			
Yes	20	27	0
No	80	73	0
Indirect tenderness:*			
Yes	68	42	+5
No	32	58	−6
Rigidity or guarding:*			
Yes	39	16	+9
No	61	84	−3
Tenderness outside the right lower quadrant:			
Yes	39	52	0
No	61	48	0
Rectal tenderness:			
Yes	50	44	0
No	50	56	0
Age (Years):			

Table I. *continued.*

Variable	Acute appendicitis (n = 94)	Normal appendix (n = 128)	Weight of score
Age (Years):			
≤10	4	11	0
11-20	38	39	0
21-65	52	43	0
≥66	6	7	0
Sex:*			
Male	67	43	+4
Female	33	57	−5

* $p < 0.05$.

acute appendicitis intensive clinical observation (7), diagnostic ultrasonography (US) (7, 10–12, 14–16, 20–22), computer or score-aided prediction (1, 2, 4–8, 13, 17, 18), laborarory tests (7), and laparoscopy (7) have been evaluated. Among these methods diagnostic laparoscopy has become an important invasive tool in the diagnosis and management of the acute abdomen. Of the remaining non-invasive diagnostic methods computer or score-aided prediction and diagnostic US have been recommended as reliable methods to avoid negative appendicectomies without increasing the risk of perforation (7).

The aim of the present study was to evaluate and compare the diagnostic accuracy of initial clinical judgment compared with initial routine diagnostic US in a prospective trial in the same group of patients admitted with suspected acute appendicitis. In addition, we have attempted to create a scoring system based on data collected prospectively.

PATIENTS AND METHODS

The study was done prospectively among 222 consecutive patients admitted with suspected acute appendicitis during a two-year period from June 1990 to June 1992 at Odense University Hospital, Denmark. Patients admitted between midnight and 0800 were excluded from the study because there were too few staff in the department of radiology. The patients ranged in aged from 4 to 98 years, (mean (SD) 27 (18) years). One hundred and four of the patients were female and 118 were male.

On admission, all patients underwent a clinical examination by the junior surgeon on call. According to clinical judgment the patients were classified as requiring immediate operation or observation. During the physical examination 21 variables relating to history, physical examination, and laboratory findings were recorded (Table I) to create the database for a scoring system. After the clinical examination, US of the lower abdomen and pelvis was done, which included detailed examination of the right iliac fossa. The study was evaluated by the local Human Use Committee and for ethical reasons it was decided that disabled patients with the suspicion of perforation, or patients for whom US examination would have caused a delay in operation, should not undergo the examination.

At the end of study, all prospectively collected data were analysed and correlated with the operative and histological findings and clinical outcome (observed group), which were considered to be the "gold-standard".

The patients were treated according to the routine of the surgical department regardless of the result of the US examination, of which the surgeon was unaware during the patient's entire hospital stay. During the study 13 junior surgeons participated, all of whom had at least two years' experience in general surgery. The state of all excised appendixes was confirmed histologically and classified as normal, acutely inflamed (diffuse infiltration of granulocytes in the tunica muscularis), gangrenous (focal areas of necrosis extending through the wall of the appendix), or perforated. Differentiation between gangrenous and perforated appendix was based on the macroscopic evaluation. In those patients who were not operated on, the diagnosis was established by evaluating all clinical findings including a 12 month follow-up.

The US examination was made immediately after the clinical examination. The radiologist was not aware of the clinical history and signs or the decision about treatment. If findings in the lower abdomen and pelvis were negative, the US examination was extended to include the entire abdomen. All US examinations were made by the staff radiologist on call, whose expertise in US ranged from doctors in training to US experienced consultants. Eighteen different radiologists took part in the study and they had at least a year's experience with the technique of US. The US scan was done with a

Table II. *Findings in 148 patients operated on for acute appendicitis*

Figures are number (%) of patients.

Condition of appendix	Sex		Total (n = 148)
	Male (n = 79)	Female (n = 69)	
Not inflamed	18 (23)	36 (52)	54 (36)
Inflamed	32 (41)	20 (29)	52 (35)
Gangrenous	19 (24)	8 (12)	27 (18)
Perforated/abscess	10 (13)	5 (7)	15 (10)

Table III. *Conditions that mimicked acute appendicitis*

Conditions	No. (%) of cases
Mesenteric adenitis	35 (65)
Colonic cancer	1 (2)
Perforated sigmoid diverticulitis	1 (2)
Gynaecological problems	5 (9)
Abdominal pain of unknown aetiology	12 (22)
Total	54

Table IV. *Tentative diagnoses in the 74 patients treated conservatively*

Diagnosis	No. (%) of cases
Non-specific gastroenteritis	45 (61)
Mesenteric adenitis	4 (5)
Cholelithiasis	4 (5)
Pancreatitis	3 (4)
Urinary tract infection or stones	8 (11)
Gynaecological disorders	10 (14)
Total	74

5 Mhz linear-array transducer (Acuson) using the graded compression technique described by Puylaert (11). Positive sonographic diagnosis of acute appendicitis was based on visualisation of a non-compressible aperistaltic tubular structure with a "bull's-eye" appearance in the transverse view with an outer diameter exceeding 6 mm, or by signs of a periappendicular abscess. Non-visualisation of the appendix with no other intra-abdominal abnormalities was recorded as a normal result. On the basis of the US findings the patient was placed into one of the three categories: appendicitis, other conditions, or normal US examination.

To create a scoring system a database was established from the 21 predictive variables recorded. Each of the predictive factors was divided into two groups according to whether the discharge diagnosis was acute appendicitis or not and tested for significance at a probability of 0.05 using a chi square test. Factors that were not significant were discarded. The diagnostic sensitivity and specificity of each of the significant predictive factors in the two groups were calculated in regard to presence and absence. The likelihood ratio was employed as a diagnostic weight, expressed as a positive or negative weight, using the following formula:

$$\text{Weight} = 10 \times \ln(\text{sensitivity}/(1-\text{specificity}))$$

All ratios were rounded to the nearest integer. Through simple addition of appropriate weights, a diagnostic score was calculated for each patient, and the possible end score varied from −75 to +54. The resulting scores were then analysed to find out the most appropriate cutoff point(s) for deciding whether to operate for appendicitis or to observe the patient.

RESULTS

Diagnostic accuracy by clinical evaluation: laparotomy was undertaken for suspected acute appendicitis in 148 (67%) of the 222 patients, of whom 94 patients had appendicitis (prevalence 42%). Among those with appendicitis, 52 had an acute appendix, 27 had a gangrenous appendix, and 15 had either a perforated appendix or a periappendiceal abscess (Table II). Of the remaining 54 patients who had laparotomy for conditions other than appendicitis, two had disorders that required an operation, 40 had other pathological intra-abdominal processes that did not require an operation, and 12 had no identifiable pathological processes (Table III).

All 74 patients (33%) who were initially selected to conservative treatment recovered from their illnesses without operation. They were all discharged after a few days with the diagnoses listed in Table IV. None of

Table V. *Correlation of operative and ultrasonographic findings in 123 patients*

Figures are number (%) of patients.

Conditions of appendix	Ultrasonographic result		Total
	Correct	Wrong	
Not inflamed	41 (91)	4 (9)	45
Inflamed	20 (45)	24 (55)	44
Gangrenous	10 (45)	12 (55)	22
Perforated/abscess	8 (67)	4 (33)	12
Total	79 (64)	44 (36)	123

these patients was readmitted with appendicitis during the following 12 months.

Diagnostic accuracy of ultrasonographic evaluation: among the 222 patients who entered the study, 25 were excluded from US examination for ethical reasons and 4 belonging to the observed group for other reasons (overall 13%). The excluded group of patients had a higher prevalence of acute appendicitis (64%) but did not differ in any other aspects being studied concerning pathology, sex, or age.

Of the remaining 193 patients (87%) who underwent US, 123 (64%) were operated on for acute appendicitis, which was confirmed in 78 (prevalence 40%). Of the confirmed cases, 38 were diagnosed by US (true positive diagnosis) and the remaining 40 were missed by US (false negative diagnosis). Among the 45 patients in whom acute appendicitis was excluded at operation, 41 had a normal US scan (true negative diagnosis) and four had a positive US scan (false positive diagnosis) (Table V).

Seventy patients were treated conservatively of whom 60 had a normal US scan (true negative diagnosis) and 10 had a positive scan (false positive diagnosis). Among the 60 patients with a normal US scan, alternative non-surgical conditions probably responsible for the abdominal symptoms were disclosed by US in 10 cases, and included mesenteric lymphadenitis ($n = 1$), cholelithiasis ($n = 4$), urological conditions ($n = 2$), and gynaecological disorders ($n = 3$).

Diagnostic accuracy by score evaluation: 11 of the 21 criteria predictive of acute appendicitis were significant ($p < 0.05$) and included: total white cell count (WCC) ($>10 \times 10^9/1$), migration of pain to the right lower quadrant, gradual onset of pain, increasing intensity of pain, pain aggravated by movement, pain aggravated by coughing, anorexia, vomiting, indirect tenderness (Rovsing's sign), muscle spasm, and sex (Table I). These 11 factors were used to create a diagnostic scoring system and through simple addition of appropriate weights a diagnostic score for each patient was calculated. Score values ranged from −75 to +54. Figure 1 shows the distribution of the diagnostic score for patients with or without appendicitis including the overlapping of diagnostic score distributions in the two groups. To analyse the information obtained from these figures two approaches were examined.

By the first approach the scoring system was evaluated as a two way score with an appropriate cutoff point for either observation or appendicectomy. With a cutoff level of "+9", calculated from the apex of the receiver operater characteristic (ROC) curve (Fig. 2), observation was suggested in 129 patients of

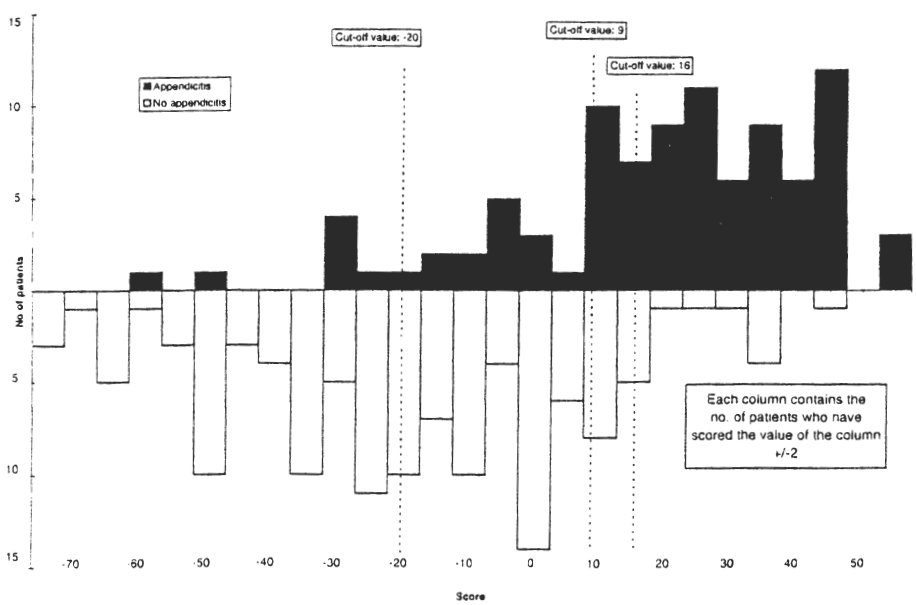

Fig. 1. Distribution of the diagnostic scores for the 94 patients with acute appendicitis, plotted over the distribution of the diagnostic scores for the 128 patients with normal appendixes.

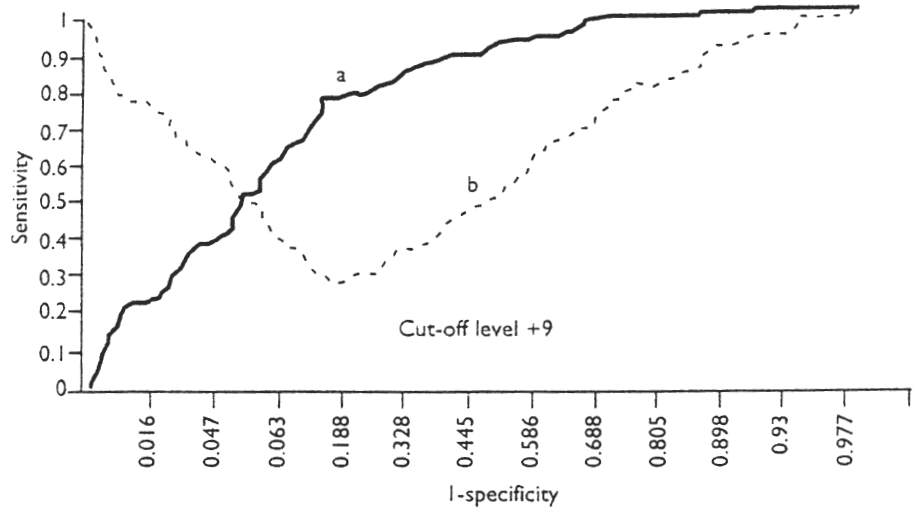

Fig. 2. a. Receiver operator characteristic (ROC) curve. *b.* Curve describing calculated deviation from theoretical best point (0.1), where sensitivity = specificity = 1 and $|D| = \sqrt{(1-x)^2 + (1-y)^2}$

whom 21 had surgically proved appendicitis (false negative). Among the 21 false negative results, 14 had an inflamed appendix, three had a gangrenous appendix, and four had a perforated appendix. Of the 93 patients in whom operation was suggested, 20 had a normal appendix (false positive).

By the second approach the score was analysed with three test zones indicating an increasing probability of appendicitis. At cutoff points at "−20" and "+16" there were three zones with low, intermediate, and high risk of having acute appendicitis. In the zone with a low risk (≤ -20) observation was suggested in 69 patients of whom seven patients had surgically proved inflamed appendixes (false negative). In the zone at high risk of appendicitis ($\geq +16$) surgery was proposed for 67 patients of whom eight patients would have had an unnecessary appendicectomy. Among the eight false positive cases, one was operated on for perforated perisigmoiditis, three were confirmed at operation, one had cholelithiasis, one had urolithiasis, and two had non-specific gastroenteritis. The intermediate zone (-19 to $+15$) included 87 patients of whom 28 had surgically proved appendicitis. Of the 28 patients with appendicitis 18 had an inflamed appendix, five had a gangrenous appendix, and five had a perforated appendix.

The combination of diagnostic US and clinical score was evaluated partly to obtain further information about the score and partly to find out whether there could be an advantage from this combination. Of the seven false negative cases in the zone at low risk, five had US scans in which one case of appendicitis was confirmed. Of the eight cases with false positive diagnoses in the zone at high risk, seven patients had diagnostic US of which one scan was positive, indicating that this patient was a true positive. Among the 28 patients with appendicitis in the intermediate zone, 22 patients had US examinations in which 12 cases of appendicitis (nine inflamed and three gangrenous or perforated) were were seen on the US scan.

DISCUSSION

Sufficient epidemiological data on the incidence of perforation in acute appendicitis are lacking and probably the most important reason for the low diagnostic accuracy in the disease. Most people accept that appendicitis might perforate, which is associated with a high morbidity and mortality, and so early exploration is normally done in suspected cases. Recently, however, there have been reports of relatively high morbidity after negative appendicectomy, which have questioned the aggressive attitude in the management of appendicitis. The primary goal should therefore be to avoid negative appendicectomies and to prevent perforation. Whether these two things can be combined or are inversely proportional is controversial.

In this study analysis of the clinical judgment yielded a diagnostic accuracy of 76%, a specificity of 58%, a negative appendicectomy rate of 36%, and a negative laparotomy rate of 35%.

Patients with equivocal symptoms of acute appendicitis or patients admitted for a second time with suspected appendicitis were operated on immediately. This reflects the observation that a ruptured appendicitis carries a morbidity much greater than that of a negative exploration (7, 18, 19) and that estimations of morbidity as high as 15% for negative appendicectomy

Table VI. *Previously reported results of ultrasonography in the diagnosis of acute appendicitis*

First author	Reference No.	Year of publication	No. of cases	Sensitivity (%)	Specificity (%)	Accuracy (%)
Puylaert	11	1986	60	89	100	95
Puylaert	12	1987	111	75	100	88
Sim	16	1989	80	90	100	92
Schwerk	14	1990	857	90	98	96
Skaane	15	1990	240	78	92	87
Vignault	20	1990	7	94	89	91
Worrell	22	1990	0	68	98	92
Wade	21	1993	110	86	84	85

have been accepted as preferable to the higher morbidity and mortality associated with a ruptured appendix. A drawback, however, was a negative appendicectomy rate considerably above the widely assumed norms (7). The incidence of perforated appendicitis were 16% (% of diseased appendixes) of which most, if not all, perforations might have occurred before referral. Diagnostic error was more than twice as common in female as in male patients (Table II) and this was not caused by gynaecological diseases because only five women had gynaecological conditions. The most common condition mimicking appendicitis among both sexes was non-specific mesenteric adenitis.

The results of clinical judgment in this study were disappointing. It is possible, however, that a more conservative management policy could have increased the diagnostic accuracy without a simultaneous increasing of the perforation rate (7). A recent regression analysis of 10000 appendicectomies done over a period of 15 years showed an inverse relationship between the normal appendicectomy rate and the rate of perforated appendicitis (19). Furthermore, the overall complication rate in patients suspected of having appendicitis improved when the rate of perforated appendicitis was lowered, even if this meant raising the negative appendicectomy rate.

The present results indicate the potential difficulty of an exact diagnosis of appendicitis and confirm the need for diagnostic aids.

To improve the diagnosis of appendicitis previous studies have evaluated the availability of ultrasonography. These studies showed sensitivities between 68% and 96%, specificities of 84% to 100%, and accuracy of 85% to 96% (Table VI). All authors agreed that US is a valuable diagnostic aid in the initial clinical evaluation of patients with suspected appendicitis. However, the various methods applied in these studies make it difficult to transfer the results of diagnostic US to the everyday surgical practice in our department. A principal charge against these studies could be that most of all US examinations were done by a small group of ultrasound-dedicated radiologists (10–12, 15, 16, 21, 22). It is well known that the accuracy of US is operator-dependent and requires dedication and experience, and the logistics of providing 24-hour radiological cover for such an US service would not be possible in most clinical departments. Secondly, in some studies the radiologist already knew what treatment had been decided (10, 14, 16, 22), and finally, one study evaluated diagnostic US solely in children (20).

The current study was designed to assess the diagnostic accuracy of immediate, routine ultrasonographic examination in the diagnosis of appendicitis. US examinations were made by the radiologist in charge at the time of admission.

If the decision to operate in this study had been based solely on US diagnosis, the result would have been a sensitivity of 49%, specificity of 88%, and accuracy of 72%. Among the patients with false positive diagnoses, most belonged to the group of observed patients and might represent "abortive appendicitis" or cases of acute appendicitis with spontaneous resolution of inflammation (10). Nevertheless, these cases were interpreted as misdiagnoses, a position which was supported by the facts that none was readmitted with symptoms of acute appendicitis. In some cases, alternative diagnoses were given by US, but none of the patients seemed to benefit from this as the initial planned management would not have been influenced by these US diagnoses.

Because of the larger number of radiologists and the different times of entrance and exit to the study it was not possible to make any valid statistical analysis of whether individual radiologists improved in diagnostic accuracy during the study.

Compared with previous US studies, our experience

with diagnostic US was less encouraging. If diagnostic US was applied to the excluded group of patients it might have increased the accuracy, but it does not affect the fact that 51% of all surgically-confirmed cases of appendicitis were overlooked. US is not meant to replace surgical judgment and should be regarded only as a diagnostic aid to obtain further clinical information. It is possible that most of the patients with false-negative US examinations would have had clinically evident appendicitis and therefore it would not have had any influence on the management. An analyse of how the US information is used and what influence it will have on the surgeon cannot be deduced from the present study.

In the light of our results we do not recommend routine diagnostic US in the diagnosis of appendicitis. We do not deny the diagnostic potential of US in detecting appendicitis, but it should be used only by ultrasound-dedicated radiologists.

Scoring systems seem to be ideal for supporting the diagnosis of appendicitis because they are accurate, non-invasive, and require no special equipment (7). Several scoring systems (1, 2, 6, 8, 13, 17) and computer analyses (4, 5, 18) have been devised to aid decision-making in the diagnosis of appendicitis. Most of the cited studies on scoring systems (2, 6, 8, 13) were evaluated clinically in a separate prospective study and two scores were retrospective (1, 17). Most authors found score and computer-aided diagnosis superior to unaided clinical diagnosis (1, 2, 4–6, 13, 17), but two studies did not observe any improvement (8, 18). However, the proponents of computer or scoring systems all emphasise that these methods should be regarded only as a diagnostic aids in decision-making and not as a substitutes for surgical judgment.

Despite the encouraging results none of these methods is in widespread use. Ohmann et al. (9) re-evaluated most of the above-mentioned predictive scores (1, 2, 6, 8, 17) in a different clinical environment from which they were constructed, and reported poor performances by all the scores. This observation suggested that a scoring system that functions well in one place may not necessarily function well in another, which makes their widespread use less likely. Furthermore, it has been suggested that a database from one centre cannot successfully be transferred to another geographical area (6, 18).

The methods applied in this scoring system were chosen partly because of their simple design and application and partly because a similar approach, converting likelihood ratios into weights, has successfully been used in previous studies (8, 19, 23).

An evaluation of the current scoring system applied in two parts yielded a diagnostic accuracy of 82%, sensitivity of 78%, and specificity of 84%. Among the false negative results, seven patients had gangrenous or perforated appendicitis, indicating intense observation for 58% of the patients and for that reason minor practicable for clinical use. If the scoring system was applied with three test zones it seemed to be of some help in supporting the selection of treatment, particularly if the score was within the groups at high and low risk of having appendicitis which made up 61% of the patients. An analysis of the group at low risk (≤ -20) showed a sensitivity of 93% and specificity of 48% indicating that this group could be suitable for observation. This assumption was further supported by the fact that none of the false negative cases had advanced stages of appendicitis.

An analysis of the group at high risk ($\geq +16$) showed a sensitivity of 63%, specificity of 94%, and an overall negative laparotomy rate of 10%. These results indicated a strong suspicion of appendicitis that may justify immediate operation.

However, in the remaining group at intermediate risk of having appendicitis there was a pronounced overlapping of diagnostic scores in the group with normal appendixes and that with appendicitis with a relative high proportion of cases with more advanced appendicitis. Within this zone the scoring system did not seem to provide any help in decision-making. How to handle these patients might depend on the surgeon's attitude and any conclusion about the benefit or risk by monitoring this group cannot be made from the present study. If the results of diagnostic US were added to the patients in the intermediate zone, it seems that nearly half of the cases of appendicitis could be seen by US. Whether this is true clinically and what influence this information would have it is not possible to analyse.

Any valid conclusion about the present score to support diagnostic decision-making cannot be made from this study and further prospective large scale trials are needed.

In conclusion, the diagnosis of acute appendicitis remains difficult, and the clinical judgment of a junior surgeon was far from optimal. Some of the explanation is the result of the present management policy. The clinical results might also give rise to the question of whether the diagnosis of appendicitis should be made solely by junior surgeons or should be reserved for more experienced surgeons. It would be desirable if the specificity could be increased without a decrease of sensitivity. Diagnostic US done by the radiologist on call seems not to fulfil this criterion, and is therefore not recommended as a routine procedure. Application of a scoring system seemed to be of some help if the score fell inside the group at either high or low risk of

having appendicitis, but any conclusion about its clinical practicability cannot be made from this study.

REFERENCES

1. Alvarado A. A practical score for the early diagnosis of acute appendicitis. Ann Emerg Med 1986; 15: 557-564.
2. 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.
3. Baigrie RJ, Dehn TCB, Fowler SM, Dunn DC. Analysis of 8651 appendicectomies in England and Wales during 1992. Br J Surg 1995; 82: 933.
4. Dombal FT, Leaper DJ, Horrocks JC, Staniland JR, McCann AP. Human and computer-aided diagnosis of abdominal pain: Further report with emphasis on performance of clinicians. B M J 1974; i: 376-380.
5. Edwards FH, Davies RS. Use of Bayesian algorithm in the computer-assisted diagnosis of appendicitis. Surg Gynecol Obstet 1984; 158: 219-222.
6. Fenyö G. Routine use of a scoring system for decision-making in suspected acute appendicitis in adults. Acta Chir Scand 1987; 153: 545-551.
7. Hoffmann J, Rasmussen OØ. Aids in the diagnosis of acute appendicitis. Br J Surg 1989; 76: 774-779.
8. Izbicki JR, Knoefel WT, Wilker DK, et al. Accurate diagnosis of acute appendicitis: A retrospective and prospective analysis of 686 patients. Eur J Surg 1992; 158: 227-231.
9. Ohmann C, Yang Q, Franke C, Abdominal Pain Study Group. Diagnostic scores for acute appendicitis. Eur J Surg 1995; 161: 273-281.
10. Ooms HWA, Koumans RKJ, Kang You PJH, Puylaert JBCM. Ultrasonography in the diagnosis of acute appendicitis. Br J Surg 1991; 78: 315-318.
11. Puylaert JBCM. Acute appendicitis: US evaluation using graded compression. Radiology 1986; 158: 355-360.
12. Puylaert JBCM, Rutgers PH, Lalisang RI, et al. A prospective study of ultrasonography in the diagnosis of appendicitis. N Engl J Med 1987; 317: 666-669.
13. Ramirez JM, Deus J. Practical score to aid decision making in doubtful cases of appendicitis. Br J Surg 1994; 81: 680-683.
14. Schwerk WB, Wichtrup B, Ruschoff J, Rothmund M. Acute and perforated appendicitis: current experience with ultrasound-aided diagnosis. World J Surg 1990; 14: 217-276.
15. Skaane P, Amland PF, Nordhus T, Solheim K. Ultrasonography in patients with suspected acute appendicitis: a prospective study. Br J Radiol 1990; 63: 787-793.
16. Sim KT, Picone S, Crade M, Sweendy JP. Ultrasound with graded compression in the evaluation of acute appendicitis. J Natl Med Assoc 1989; 81: 954-957.
17. Teicher I, Landa B, Cohen M, Kabnick LS, Wise L. Scoring system to aid in the diagnosis of appendicitis. Ann Surg 1983; 198: 753-759.
18. Van Way CW, III, Murphy JR, Dunn EL, Elerding SC. A feasibility study of computer aided diagnosis in appendicitis. Surg Gynecol Obstet 1982; 155: 685-688.
19. Velanovich V, Satava R. Balancing the normal appendectomy rate with the perforated appendicitis rate: implications for quality assurance. Am Surg 1992; 58: 264-269.
20. Vignault F, Filiatrault D, Brandt ML, Garel L, Grignon A, Ouime A. Acute appendicitis in children: evaluation with US. Radiology 1990; 176: 501-504.
21. Wade DR, Morrow SE, Balsara ZN, Burkhard TK, Goff WB. Accuracy of ultrasound in the diagnosis of acute appendicitis compared with the surgeon's clinical impression. Arch Surg 1993; 128: 1039-1046.
22. Worell JA, Drolshagen LF, Kelly TC, Hunton DW, Durmon GR, Fleischer AC. Graded compression ultrasound in the diagnosis of appendicitis. J Ultrasound Med 1990; 9: 145-150.

RÉSUMÉ

But: Evaluer la précision diagnostique de l'examen clinique et de l'échographie faite à titre systématique, et établir un score d'aide au diagnostic.
Type d'étude: Prospective, en double aveugle.
Provenance: Hôpital universitaire, Danemark.
Patients: Deux cent vingt-deux patients consécutifs admis entre 8 heures et minuit pour suspicion d'appendicite de Juin 1990 à Juin 1992.
Méthodes: Cent quarante-huit patients (67%) ont eu une appendicectomie et les 74 autres ont été placés en observation. Cent quatre-vingt-treize patients (87%) ont eu une échographie; Vingt et une variables prédictives ont été recueillies de façon prospective pour établir un score.
Principaux critères de jugement: Les résultats de l'examen anatomopathologique, l'évolution clinique (chez les patients placés en observation), les données de l'échographie, et les valeurs du score diagnostique.
Résultats: L'indication opératoire a été portée par un jeune chirurgien sur les seules données de l'examen clinique, avec une précision diagnostique de 76%, une spécificité de 58%, et un taux d'appendicectomie avec appendice sain de 36%. Cent quatre-vingt- treize patients ont eu une échographie faite par le radiologue de garde, et parmi ceux-ci, 123 ont été opérés d'une appendicite histologiquement prouvée. L'échographie avait une précision diagnostique de 72%, une sensibilité de 49%, et une spécificité de 88%. Parmi les 21 facteurs prédictifs d'appendicite aiguë, 11 avaient une valeur significative ($p < 0{,}05$): un chiffre de globules blancs supérieur à 10×10^9/ml, la migration de la douleur dans la fosse iliaque droite, son début progressif, son intensité croissante, le fait qu'elle soit exacerbée par les mouvements et la toux, l'existence d'une anorexie et de vomissements, une douleur à la décompression (signe de Rovsing), une défense et le sexe. Une importance plus ou moins grande a été accordée à ces 11 facteurs selon leur taux de probabilité, et a été utilisée pour établir un score. Ce score n'était pas très performant lorsqu'il était utilisé pour distinguer les patients placés en observation de ceux opérés. Cependant, s'il était utilisé avec deux points de séparation délimitant trois zones (risque faible, intermédiaire et élevé d'appendicite aiguë) il apparaissait un certain bénéfice diagnostique chez les patients se situant dans les zones de faible risque et de risque élevé.
Conclusions: L'impression clinique des jeunes chirurgiens est insuffisante et des aides au diagnostic sont nécessaires pour diminuer le taux des appendicectomies inutiles. L'échographie faite à titre systématique n'est pas performante. Le recours à un nouveau score diagnostique pourrait apporter une certaine aide chez les patients ayant un risque faible ou élevé d'avoir une appendicite aiguë mais nous ne

pouvons tirer aucune conclusion quant à son application en pratique clinique.
Mots-clés: Appendicite aiguë, appendicectomie, précision diagnostique, échographie à visée diagnostique, score diagnostique, système d'évaluation, étude prospective.

ZUSAMMENFASSUNG

Ziel: Die Evaluierung der diagnostischen Genauigkeit der klinischen Beurteilung und der routinemäßig durchgeführten diagnostischen Ultrasonographie sowie die Entwicklung eines Scoringsystems zur Unterstützung der Diagnose.
Studienanordnung: Prospektive Doppelblindstudie.
Studienort: Universitätskrankenhaus, Dänemark.
Patienten: 222 konsekutive Patienten mit vermuteter akuter Appendizitis, aufgenommen zwischen 8.00 Uhr and 24.00 Uhr von Juni 1990 bis Juni 1992.
Methoden: 148 Patienten (67%), die eine Appendektomie erhielten und die verbleibenden 74 Patienten wurden beobachtet. 193 (87%) erhielten eine diagnostische Ultrasonographie-Untersuchung. 21 prediktive Variablen wurden prospektiv gesammelt zur Entwicklung eines Scoringsystems.
Endpunkte: Ergebnisse der chirurgisch-pathologischen Untersuchungen, klinisches Outcome (Beobachtungsgruppe), diagnostische Ultrasonographie, und der Wert des diagnostischen Scores.
Ergebnisse: Die Entscheidung zur Operation wurde von einem Juniorchirurgen selbständig vorgenommen auf der Grundlage der klinischen Untersuchung, die eine diagnostische Genauigkeit von 76%, Spezifität von 58% und negative Appendektomierate von 36% aufwies. 193 Patienten erhielten eine diagnostische Ultrasonographie-Untersuchung, die durch den diensthabenden Radiologen durchgeführt wurde, von denen 123 Patienten operiert wurden, bei denen wiederum in 78 Fällen ein histologischer Nachweis einer Appendizitis gegeben war. Ultrasonographie zeigte eine diagnostische Genauigkeit von 72%, Sensitivität von 49% und Spezifität von 88%. Von den 21 prediktiven Faktoren für die akute Appendizitis zeigten sich 11 als signifikant ($p < 0{,}05$): Leukozytenzahl, Migration des Schmerzes zum rechten unteren Quadranten, graduelles Einsetzen von Schmerzen, Zunahme der Schmerzintensität, Schmerzzunahme bei Bewegung, Schmerzzunahme bei Husten, Anorexia, Erbrechen, indirekte Abwehr (Rovsingzeichen), Muskelspasmus, Geschlecht. Diese 11 Prediktoren wurden entsprechend auf der Basis des Wahrscheinlichkeitsverhältnisses gewichtet und zur Entwicklung eines Scoringsystems herangezogen. Der Score schnitt schlecht ab in der Beurteilung der Patienten zur Beobachtung oder Appendektomie. Wenn er jedoch mit 2 Cut-off-Punkten und damit 3 Testzonen (Niedrig-, Intermediär- und Hochrisiko einer akuten Appendizitis) verwendet wurde, zeigte sich ein diagnostischer Vorteil für die Patienten in den Zonen der niedrigen und hohen Wahrscheinlichkeit.
Schlußfolgerungen: Die klinische Beurteilung des Juniorchirurgen war enttäuschend und diagnostische Mittel sind wünschenswert, um die negative Appendektomierate zu reduzieren. Die diagnostische Ultrasonographie erwies sich als schwache Routinemethode. Die Applikation eines Scoringsystems könnte bei manchen Patienten der niedrigen oder hohen Wahrscheinlichkeit einer akuten Appendizitis hilfreich sein, doch wir können zu keiner Schlußfolgerung über seine klinische Applikation kommen.
Schlüsselwörter: Akute Appendizitis, Appendektomie, diagnostische Genauigkeit, sonographische Diagnose, Ultraschall, Untersuchung, diagnostisches Score, Scoringsystem, prospektive Studie.

РЕЗГОМЁ

Цель: Изучить диагностическую точность клинического обследования и ультрасонографии у пациентов с острым аппендицитом а также разработка скоринг-системы для диагностики острого аппендицита.
Характер исследования: Проспективное исследование методом двойного слепого контроля.
Клиника: Университетский госпиталь, Дания.
Пациенты: 222 консекутивных пациента с подозрением на острый аппендицит, леченных за период времени с июня 1990 по июнь 1992.
Методы: 148 пациентов (67%) подверглись аппендектомии, остальные 74 клиническому наблюдению. У 193 пациентов (87%) было выполнено ультразвуковое исследование органов брюшной полости. С целью развития диагностической скоринг-системы был изучен 21 диагностический фактор.
Задачи исследования: Изучение результатов хирургических находок, патологоанатомического исследования, клинических данных (в группе наблюдения), а также изучение данных ультразвукового исследования и оценка данных диагностической скоринг-системы.
Результьаты: Решение о выполнении неотложной операции принималось мододыми хирургами только на основании клинического исследования, причем точность клинического обследования составила 76%, специфичность 58% и в 36% случаев была произведена негативная аппендэктомия. У 193 пациентов было выполнено ультразвуковое исследование органов брюшной полости, которое производилось сотрудниками радиалогической клиники. Из 123 произведенных аппендэктомий в 78 случаях был гистологически подтвержден острый аппендицит. Точность ультразвуковой диагностики острого аппендицита составляла 72%, чувствительность 49% и специфичность 88%. Из 21 диагностических факторов острого аппендицита 11 были статистически достоверны ($p < 0.05$). К этим факторам относятся лейкоцитоз, миграция боли в правую подвздошную область, постепенное начало болей, возрастание интенсивности боли, усиление болей при движении, кашле,рвоте и тошноте, положительный симптом Ровзинга, напряжение мышц передней брюшной стенки и пол. Зти 11 факторов были разполодены в соответствии с их значимостью и использованы для создания скоринговой системы. Данные скоринговой системы были недостаточны чтобы отделить пациентов, нуждающихся в неотложной операции от пациентов, нуждающихся в наблюдении. Однако, если эта скоринговая система была разделена на з зоны (зона с низкой, зона со средней и зона с высокой вероятностью острого аппендицита) тогда большая вероятность правидьного диагноза наблюдалась у пациентов внутри зоны с низкой и высокой вероятностью острого аппендицита.
Выводы: Клиническая диагностика, производимая

молодыми хирургами, была недостаточна для правильной диагностики аппендицита, поэтому желательна дополнительная диагностическая помощь с целью уменьшения количества негативных аппендэктомий. Полученные данные показали плохую выполнимость ультразвукового исследования в качестве рутинной процедуры. Использование полученной скоринг-системы может оказать некоторую помощь в различии пациентов с высокой и низкой вероятностью острого аппендицита, однако данная скоринг-система не может быть рекомендована для практического применения в клинике.

Ключевые слова: Острый аппендицит, аппендэктомия, диагностическая точность, сонографическая диагностика, ультразвук, диагностическая скоринг-система, диагностический скоринг, проспективное исследование.

Submitted August 16, 1995; submitted after revision April 12, 1996; accepted April 17, 1996

Address for correspondence:
Henrik Jahn, M.D.
Department of Surgery
Odense University Hospital
DK-5000 Odense C
Denmark