

Meta-analysis of the clinical and laboratory diagnosis of appendicitis

R. E. B. Andersson

Department of Surgery, County Hospital Ryhov, SE-551 85 Jönköping, Sweden
Correspondence to: Dr R. E. B. Andersson (e-mail: roland.andersson@ltjkg.se)

Background: The importance of specific elements in the clinical diagnosis of appendicitis is controversial. This review analyses the diagnostic value of elements of disease history, clinical findings and laboratory test results in suspected appendicitis.

Methods: A systematic Medline search was made of all published studies on the clinical and laboratory diagnosis of appendicitis in patients admitted to hospital with suspected disease. Meta-analyses of receiver–operator characteristic (ROC) areas, and positive and negative likelihood ratios, of 28 diagnostic variables described in 24 studies are presented.

Results: Inflammatory response variables (granulocyte count, proportion of polymorphonuclear blood cells, white blood cell count and C-reactive protein concentration), descriptors of peritoneal irritation (rebound and percussion tenderness, guarding and rigidity) and migration of pain were the strongest discriminators, with ROC areas of 0.78 to 0.68. The discriminatory power of the inflammatory variables was particularly strong for perforated appendicitis, with ROC areas of 0.85 to 0.87. Appendicitis was likely when two or more inflammatory variables were increased and unlikely when all were normal.

Conclusion: Although all clinical and laboratory variables are weak discriminators individually, they achieve a high discriminatory power when combined. Laboratory examination of the inflammatory response, clinical descriptors of peritoneal irritation, and a history of migration of pain yield the most important diagnostic information and should be included in any diagnostic assessment.

Paper accepted 11 November 2003

Published online in Wiley InterScience (www.bjs.co.uk). DOI: 10.1002/bjs.4464

Introduction

The decision to explore a patient with suspected appendicitis is based mainly on the disease history and the findings at physical examination. The clinical presentation is, however, seldom typical and diagnostic errors are common^{1–3}. New diagnostic techniques, such as computer-aided diagnosis, diagnostic scoring systems, peritoneal aspiration cytology, diagnostic laparoscopy, leucocyte scintigraphy, ultrasonography and computed tomography, are expensive and not widely available at all times of the day and night. Furthermore, their reportedly good results have not always been reproduced under everyday conditions. Clinical diagnosis therefore remains the cornerstone of management⁴. A thorough clinical examination, including rectal palpation, is often stressed as an essential part of diagnosis, with laboratory examinations commonly thought of as less important because of their putative discriminatory capacity^{2,5–8}.

However, the diagnostic usefulness of the clinical signs has seldom been assessed and compared with that of their laboratory counterparts.

The diagnostic value of clinical signs, symptoms and laboratory tests is influenced by the incidence of the disease under study and by the spectrum of diseases in the study population⁹. For instance, the diagnostic value of C-reactive protein (CRP) concentration in appendicitis is much greater for patients admitted to hospital with suspected appendicitis than for those submitted to surgery because of clinically suspected appendicitis¹⁰. The selection of the study population is, therefore, important when assessing the diagnostic value of any variable. A variable's diagnostic performance is best evaluated in the population on which it is intended to be used. The most crucial point in the management of patients with appendicitis is the decision about which patients need operation. The appropriate population for the study of the clinical

diagnosis of appendicitis is, therefore, patients admitted to hospital with a clinical suspicion of the disease. This review examines the diagnostic value of elements of disease history, symptoms, clinical signs and simple laboratory tests, as obtained from studies of patients admitted to hospital with suspected appendicitis.

Patients and methods

A Medline search was conducted on 1 June 2003 using the following strategy: "Appendicitis"[MESH] AND ("Signs and Symptoms"[MESH] OR "Physical examination"[MESH] OR "Laboratory Techniques and Procedures"[MESH] OR "Diagnostic Tests, Routine"[MESH] OR "Medical History Taking"[MESH] OR "Leukocytes"[MESH] OR "Acute-Phase Proteins"[MESH] OR "Skin temperature"[MESH] OR "Sensitivity and Specificity"[MESH]) NOT "Letter"[Publication Type] NOT "Case Report"[MESH]. Along with the Medline search, the list of references in identified studies and the author's own personal file maintained over 10 years of constant screening of the literature were searched for additional eligible studies.

Only studies based on patients admitted to hospital for suspected appendicitis were eligible. Studies based on unselected patients with abdominal pain, patients submitted to surgery for suspected appendicitis, or comparisons of patients with verified appendicitis with healthy controls were not eligible. Only results presented in such a way that the likelihood ratios and/or areas under the receiver-operator characteristic (ROC) curve could be calculated were included. Papers written in English, German, French, Italian, Spanish, Portuguese or the Scandinavian languages were considered for inclusion; those that included only children were not.

Excluded studies

The Medline search yielded 1728 citations. The majority of these had no relevance to the clinical diagnosis of appendicitis, or were reviews of other studies. About 240 studies presented original data relating to the clinical diagnosis of appendicitis. Fifty-four studies of the diagnosis of appendicitis in patients admitted for suspected appendicitis were identified; 30 of these were not selected for analysis because they included only children¹¹⁻²², the results were presented in a way that did not permit extraction of data for analysis^{16,23-32}, the presented results were inconsistent^{33,34}, patients with other diagnoses found at exploration (but not requiring surgical treatment) were included among those with appendicitis³⁵, non-consecutive

patients were selected at convenience^{36,37} or unusual variables were included that were not analysed by any other study (caecal squelch, continuous tenderness on prolonged pressure over McBurney's point and the value of repeat examination)³⁸⁻⁴⁰.

'Gold standard' and quality criteria

Histopathological diagnosis is usually regarded as the 'gold standard', but this is an imperfect situation for two reasons. First, there are no universally accepted criteria for the histopathological diagnosis of appendicitis and, second, histopathological examination of the appendix is not available for non-operated patients.

Mucosal inflammation, which is seen in up to 30 per cent of patients without appendicitis, is accepted by some as a sign of early disease, but others require a transmural inflammation for the diagnosis⁴¹. Non-operated patients who do not develop appendicitis during follow-up are assumed not to have had appendicitis, but epidemiological studies, and ultrasonography and computed tomography, suggest that spontaneous resolution is common⁴²⁻⁴⁴. Differences in the range of indications for surgical exploration and in the histopathological criteria for appendicitis may, therefore, influence the number of patients with resolving appendicitis who are detected, and the number of patients without appendicitis who are misclassified because of minor microscopic findings. This may lead to seriously biased results. Advanced appendicitis with gangrene or perforation necessitates more urgent surgical treatment, has a clearer definition and is less likely to be undetected because of spontaneous resolution. The diagnostic performance of a variable in advanced appendicitis is therefore less likely to be biased.

Criteria for the histopathological diagnosis and the range of indications for surgery, as shown by the proportion of negative explorations and perforations, should be considered in the evaluation of study results. The diagnostic performance of a variable in advanced appendicitis should be presented separately.

Extraction of data

Information extracted from articles was used to construct two by two tables describing the presence of the diagnosis of appendicitis and the variable. For studies that presented data with more than one cutoff point, the corresponding two by 'n' tables were extracted. For studies that presented only sensitivity and specificity, the required values were calculated from these figures and from the total number of

patients with and without appendicitis. For some studies the data were extracted from diagrams.

Outcome variables

Areas under the ROC curves and likelihood ratios were used to describe the diagnostic performance of a variable. These measures are better suited to describing and comparing the performance of weak diagnostic variables than sensitivity and specificity. The ROC curve is the plot of the proportions of true positive *versus* true negative results for each value of the study variable. The area under this curve represents the variable's discriminatory power. An area of 0.50 represents a variable with no discriminatory capacity, and an area of 1.00 indicates a perfect discriminator. For ordinal or continuous variables all the presented data were used to calculate the area. The area under the ROC curve was calculated using a non-parametric method for each variable and study⁴⁵.

Likelihood ratios express the predictive power of a variable; this may be high even when the overall discriminatory power is low. The likelihood ratio of a positive test result (LR^+) indicates how much a positive result will increase the odds of appendicitis compared

with the pretest odds. The likelihood ratio of a negative test result (LR^-) indicates the decrease in the odds of appendicitis that is associated with a negative result. Normally a LR^+ of more than 10 and a LR^- of less than 0.01 are required for a true diagnostic test.

When appropriate, the likelihood ratios were calculated for different cutoff points. Some studies present the results of a combination of variables. When the variables are combined with Boolean 'AND' and 'OR', it is possible to define three diagnostic levels representing none, at least one, or all of the variables present. For these situations multilevel likelihood ratios were calculated⁴⁶. Multilevel likelihood ratios describe the change in the odds of disease at the observed level of the test result.

The weighted pooled estimates of the ROC areas and the likelihood ratios were calculated. The confidence intervals were calculated according to a fixed or a random-effects model, depending on the result of the homogeneity test. Only data with the same definitions or cutoff points were pooled, except for some cases where the difference was judged to have no clinical importance. Some studies looked at variables that were not analysed in any other study; this review was limited to variables that were presented in more than one study.

Table 1 Studies analysing the clinical and laboratory diagnosis of appendicitis in patients with suspected appendicitis

Reference	Year	No. of patients	Appendicitis	Perforation*	Negative exploration†
Albu <i>et al.</i> ⁴⁷	1994	56	26 (46.4)	6 (23.1)	4 of 30 (13.3)
Alshehri <i>et al.</i> ⁴⁸	1995	123	66 (53.7)	—	—
Andersson <i>et al.</i> ⁴⁹	1999	496	194 (39.1)	28 (14.4)	59 of 302 (19.5)
Björkstén and Wählby ⁵⁰	1974	95	44 (46.3)	20 (45.5)‡	17 of 51 (33.3)
Bolton <i>et al.</i> ⁵¹	1975	100	46 (46.0)	8 (17.4)	5 of 54 (9.3)
Davies <i>et al.</i> ⁵²	1991	60	31 (51.7)	—	6 of 29 (20.7)
Davies <i>et al.</i> ⁵³	1993	100	42 (42.0)	—	18 of 58 (31.0)
Dixon <i>et al.</i> ⁵⁴	1991	1204	449 (37.3)	—	—
Dueholm <i>et al.</i> ⁵⁵	1989	204	59 (28.9)	8 (13.6)	38 of 145 (26.2)
Emery <i>et al.</i> ⁵⁶	1994	86	23 (26.7)	—	8 of 63 (12.7)
Erkasap <i>et al.</i> ⁵⁷	2000	102	49 (48.0)	10 (20.4)	6 of 53 (11.3)
Fenyö <i>et al.</i> ⁵⁸	1997	1167	392 (33.6)	65 (16.6)	83 of 775 (10.7)
Gil <i>et al.</i> ⁵⁹	2000	109	32 (29.4)	3 (9.4)	4 of 77 (5.2)
Golledge <i>et al.</i> ⁶⁰	1996	100	44 (44.0)	—	14 of 56 (25.0)
Hallan <i>et al.</i> ^{61,62}	1997	257	98 (38.1)	25 (25.5)	22 of 159 (13.8)
Izbicki <i>et al.</i> ⁶³	1992	150	54 (36.0)	2 (3.7)	26 of 96 (27.1)
Jahn <i>et al.</i> ⁶⁴	1997	222	94 (42.3)	15 (16.0)	54 of 128 (42.2)
John <i>et al.</i> ⁶⁵	1993	111	55 (49.5)	11 (20.0)	8 of 56 (14.3)
Middleton <i>et al.</i> ⁶⁶	1996	118	63 (53.4)	—	16 of 55 (29.1)
Miskowiak and Burcharth ⁶⁷	1982	312	100 (32.1)	—	67 of 212 (31.6)
Raftery ⁶⁸	1976	175	106 (60.6)	—	26 of 69 (37.7)
Søndenaa <i>et al.</i> ⁶⁹	1992	158	62 (39.2)	—	—
Thimsen <i>et al.</i> ⁷⁰	1989	70	28 (40.0)	8 (28.6)	—
van Dieijen-Visser <i>et al.</i> ³²	1991	258	69 (26.7)	13 (18.8)	—

Values in parentheses are percentages. *Proportion of patients with appendicitis with perforation; †proportion of patients without appendicitis submitted to exploration. ‡Advanced appendicitis (perforated or gangrenous appendicitis).

Results

Quality of included studies

Twenty-four eligible studies were selected for the analyses (Table 1). The proportion of patients with a diagnosis of appendicitis ranged from 26.7 to 60.6 (median 41.0) per cent. The proportion of appendicitis with perforation was reported in 13 studies, and ranged from 3.7 to 28.6 (median 17.4) per cent. The proportion of patients without

appendicitis who had been submitted to appendicectomy was reported in 19 studies, and ranged from 5.2 to 42.2 (median 20.1) per cent. The majority of the studies were prospective and included consecutive patients, but one had collected the data retrospectively⁵², six included non-consecutive patients or excluded patients owing to missing values^{37,48,58,61-64}, and two included only patients who had had symptoms for more than 12 h^{47,70}.

Table 2 Discriminatory power of variables described in more than one study, expressed as pooled receiver–operator characteristic area

	References	No. of patients	No. with appendicitis	Pooled ROC area	P*
Patient details and disease history					
Age	49, 64	718	288	0.58 (0.54, 0.63)	0.260
Male sex	49, 58, 61–64	2292	832	0.62 (0.60, 0.64)	0.663
Duration of symptoms	49, 58, 59, 63, 64	2141	764	0.58 (0.53, 0.63)	< 0.001
History of fever	49, 60	570	231	0.60 (0.55, 0.64)	0.028
Symptoms					
Gastrointestinal dysfunction					
Anorexia	60–62, 64	579	236	0.58 (0.54, 0.63)	0.831
Nausea	64, 65	333	149	0.56 (0.50, 0.62)	0.399
Nausea or vomiting	59, 61, 62	366	130	0.54 (0.48, 0.60)	0.843
Vomiting	49, 58, 60, 64	1965	715	0.59 (0.56, 0.62)	0.749
Pain					
Migration of pain	49, 58, 60–62, 64, 65	2459	912	0.68 (0.63, 0.74)	< 0.001
Progression of pain	58, 64	1389	486	0.61 (0.58, 0.64)	0.123
Peritonism					
Aggravation by cough	58, 60–62, 64	1746	628	0.64 (0.61, 0.66)	0.881
Aggravation by movement	49, 60–62, 64	949	418	0.59 (0.56, 0.62)	0.374
Signs					
Tenderness					
Presence or intensity	49, 54, 61–63	1928	739	0.62 (0.48, 0.76)	< 0.001
Localization	49, 58, 60, 64	1982	723	0.60 (0.55, 0.65)	0.030
Indirect tenderness	48, 49, 63, 64	988	397	0.65 (0.55, 0.75)	< 0.001
Rectal tenderness	49, 54, 59, 64, 68	1951	763	0.51 (0.48, 0.54)	0.155
Psoas sign	63, 65	261	109	0.58 (0.51, 0.65)	0.555
Peritonism					
Rebound	48, 49, 54, 58, 60–63, 65	3439	1287	0.70 (0.65, 0.75)	< 0.001
Percussion	64, 65	211	99	0.70 (0.63, 0.78)	0.793
Guarding	48, 49, 54, 60, 63, 65	2004	906	0.68 (0.60, 0.76)	< 0.001
Guarding or rigidity	61, 62, 64	479	192	0.63 (0.57, 0.68)	0.819
Rigidity	48, 54, 58	2310	841	0.57 (0.50, 0.64)	0.001
Laboratory tests and fever					
WBC	49–52, 55, 58, 59, 61–64, 66–68	3382	1288	0.77 (0.75, 0.78)	0.171
Granulocyte count	48–50	628	254	0.78 (0.75, 0.82)	0.370
Proportion of PMN cells	49, 55, 61, 62, 69	1067	427	0.77 (0.70, 0.84)	< 0.001
CRP level	47, 49, 52, 55, 57, 61, 62, 69, 70	1360	529	0.75 (0.66, 0.85)	< 0.001
Body temperature	49, 64	714	86	0.64 (0.52, 0.77)	0.005
Axillary/rectal temperature difference	59, 63	230	76	0.51 (0.43, 0.59)	0.478
RLQ/LLQ temperature difference	53, 56, 67	304	128	0.55 (0.48, 0.61)	0.414
Perforated appendicitis					
WBC	49–51, 57	614	159	0.85 (0.81, 0.89)	0.080
Granulocyte count	49, 50	399	82	0.86 (0.81, 0.90)	0.714
CRP level	49, 57	434	87	0.87 (0.74, 1.01)	< 0.001

Values in parentheses are 95 per cent confidence intervals. ROC, receiver–operator characteristic; WBC, white blood cell count; PMN, polymorphonuclear; CRP, C-reactive protein; RLQ, right lower quadrant; LLQ, left lower quadrant. *Heterogeneity test.

The diagnosis was confirmed by histopathological examination in all studies except one³⁷. Four studies provided information about the histopathological criteria for diagnosis, with granulocyte invasion in the mucosa⁵⁸, in the tunica muscularis^{57,64} or transmurally⁴⁹ as criteria for appendicitis. In one study the histological findings were reviewed by a pathologist who was blinded to the surgeon's preoperative diagnosis and the initial pathologist's report⁴⁹. The diagnosis of 'not appendicitis' in unoperated patients was supported by follow-up after discharge in five studies^{49,60,63-65}. In five studies the surgeons were blinded to the results of the laboratory tests or temperature reading^{47,53,55,56,67}. Four studies presented results for the diagnosis of advanced appendicitis separately^{49,50,51,57}. The data were partly extracted from diagrams in two studies^{51,67}.

Discriminatory power

The discriminatory power of the variables, expressed as ROC areas, is presented in *Table 2*. Inflammatory response variables (granulocyte count, proportion of polymorphonuclear (PMN) blood cells, white blood cell count (WBC) and CRP concentration) appeared to be the strongest discriminators, with a ROC area of 0.78 to 0.75, followed by the descriptors of peritoneal irritation (rebound, percussion tenderness and guarding) and migration of pain, with a ROC area of 0.70 to 0.68. The same relative order of importance of WBC and rebound tenderness was found in the four studies that analysed these two variables in identical populations, suggesting that this result was not due to the pooling of disparate studies from different study populations. A similar result was obtained when only studies of higher quality were selected

Table 3 Predictive power of elements of history and clinical examination in the diagnosis of appendicitis, expressed as pooled likelihood ratios

	LR ⁺	P*	LR ⁻	P*
Patient details and disease history				
Age ≥ 20 years	1.25 (1.10, 1.42)	0.505	0.74 (0.62, 0.89)	0.303
Male sex	1.62 (1.49, 1.76)	0.620	0.62 (0.57, 0.68)	0.340
Duration (h)				
> 9	1.01 (0.97, 1.05)	1.000	0.94 (0.62, 1.42)	0.634
> 12	0.96 (0.90, 1.04)	0.094	1.19 (0.87, 1.63)	0.107
> 24	0.65 (0.47, 0.90)	0.002	1.47 (1.14, 1.90)	< 0.001
> 48	0.49 (0.36, 0.67)	0.144	1.20 (1.08, 1.34)	0.018
History of fever	1.64 (0.89, 3.01)	0.008	0.61 (0.49, 0.77)	0.089
Symptoms				
Gastrointestinal dysfunction				
Anorexia	1.27 (1.14, 1.41)	0.927	0.59 (0.45, 0.77)	0.321
Nausea or vomiting	1.15 (1.04, 1.36)	0.657	0.72 (0.57, 0.91)	0.823
Vomiting	1.63 (1.45, 1.84)	0.455	0.75 (0.69, 0.80)	0.687
Pain				
Pain migration	2.06 (1.63, 2.60)	< 0.001	0.52 (0.40, 0.69)	< 0.001
Pain progression	1.39 (1.29, 1.50)	0.097	0.46 (0.27, 0.77)	0.043
Peritonism				
Aggravation by cough	1.49 (1.40, 1.59)	0.711	0.38 (0.32, 0.46)	0.536
Aggravation by movements	1.24 (1.16, 1.33)	0.070	0.49 (0.39, 0.62)	0.565
Signs				
Tenderness				
Direct tenderness	1.29 (1.06, 1.57)	< 0.001	0.25 (0.12, 0.53)	0.006
Indirect tenderness	2.47 (1.38, 4.43)	< 0.001	0.71 (0.65, 0.77)	0.082
Localized <i>versus</i> diffuse tenderness	1.52 (1.21, 1.92)	0.016	0.67 (0.61, 0.75)	0.760
Rectal tenderness	1.03 (0.83, 1.27)	0.043	0.96 (0.85, 1.08)	0.037
Psoas sign	2.31 (1.36, 3.91)	0.195	0.85 (0.76, 0.95)	0.243
Peritonism				
Rebound tenderness	1.99 (1.61, 2.45)	< 0.001	0.39 (0.32, 0.48)	0.004
Percussion tenderness	2.86 (1.95, 4.21)	0.244	0.49 (0.37, 0.63)	0.820
Guarding	2.48 (1.60, 3.84)	< 0.001	0.57 (0.48, 0.68)	0.015
Guarding or rigidity	2.36 (1.76, 3.15)	0.721	0.70 (0.61, 0.80)	0.605
Rigidity	2.96 (2.43, 3.59)	0.220	0.86 (0.72, 1.02)	< 0.001

Values in parentheses are 95 per cent confidence intervals. LR, likelihood ratio. *Heterogeneity test.

for analysis. The other variables had weak discriminatory power, with ROC areas of less than 0.65. Rectal tenderness had no discriminatory power (ROC area 0.51).

Predictive power

The predictive power of the variables, expressed as likelihood ratios, are presented in *Tables 3* and *4*. Appendicitis was more likely in patients with a strong inflammatory response, high granulocyte count or WBC, high proportion of PMN cells or increased CRP concentration, and a LR⁺ of 7.09 to 2.39. The signs of peritoneal irritation (rigidity, percussion and rebound tenderness) were also important predictors, with a LR⁺ of 2.96 to 1.99. Appendicitis was unlikely when an inflammatory response or peritoneal irritation was absent, as shown by a LR⁻ of 0.24 to 0.39 at a low proportion of PMN cells, at a low WBC or

granulocyte count, when the CRP concentration was normal or when direct tenderness or rebound tenderness was absent.

Diagnostic value in perforated appendicitis

Four studies presented the results for perforated appendicitis separately (*Tables 2* and *4*)^{49-51,57}. The WBC and granulocyte count, and the CRP level, had a stronger discriminatory capacity for perforated appendicitis (ROC area of 0.85 to 0.87 *versus* 0.78 to 0.75 for appendicitis). High WBC and granulocyte count, and an increased CRP concentration, were relatively strong predictors of perforated appendicitis, with a LR⁺ of up to 7.20. Perforated appendicitis was very unlikely in patients with a low WBC and granulocyte count, and a CRP concentration of less than 10 g/l, with a LR⁻ of 0.20 to 0.11.

Table 4 Predictive power of laboratory variables and body temperature in the diagnosis of appendicitis, expressed as pooled likelihood ratios

	LR ⁺	P*	LR ⁻	P*
Laboratory tests and fever				
WBC ($\times 10^9/l$)				
≥ 10	2.47 (2.06, 2.95)	< 0.001	0.26 (0.18, 0.36)	< 0.001
≥ 12	2.75 (1.99, 3.80)	0.041	0.48 (0.41, 0.55)	0.215
≥ 14	2.96 (2.48, 3.53)	0.945	0.69 (0.55, 0.86)	< 0.001
≥ 15	3.47 (1.55, 7.77)	0.012	0.81 (0.69, 0.95)	0.008
Granulocyte count ($\times 10^9/l$)				
≥ 7	1.64 (0.87, 3.09)	< 0.001	0.31 (0.23, 0.40)	0.670
≥ 9	2.66 (1.39, 5.09)	0.015	0.45 (0.37, 0.54)	0.094
≥ 11	4.36 (2.83, 6.73)	0.085	0.60 (0.53, 0.69)	0.154
≥ 13	7.09 (4.06, 12.37)	0.328	0.74 (0.68, 0.81)	0.277
Proportion of PMN cells (%)				
> 75	2.44 (1.60, 3.74)	0.001	0.24 (0.11, 0.50)	< 0.001
> 85	3.82 (2.86, 5.08)	0.158	0.58 (0.51, 0.66)	0.166
CRP level (mg/l)				
> 10	1.97 (1.58, 2.45)	< 0.001	0.32 (0.20, 0.51)	< 0.001
> 20	2.39 (1.67, 3.41)	0.042	0.47 (0.28, 0.81)	0.001
Body temperature (°C)				
> 37.7	1.57 (0.90, 2.76)	0.002	0.65 (0.31, 1.36)	< 0.001
> 38.5	1.87 (0.66, 5.32)	0.023	0.89 (0.71, 1.12)	< 0.001
Temperature difference (> 1°C)				
Axillary/rectal	1.10 (0.61, 1.96)	0.083	0.96 (0.84, 1.10)	0.282
RLQ/LLQ	1.99 (1.08, 3.86)	0.317	0.91 (0.83, 1.01)	0.123
Perforated appendicitis				
WBC ($\times 10^9/l$)				
≥ 10	4.20 (2.11, 8.35)	0.005	0.20 (0.10, 0.41)	0.082
≥ 15	7.20 (4.31, 12.00)	0.317	0.66 (0.56, 0.78)	0.595
Granulocyte count ($\times 10^9/l$)				
≥ 7	2.89 (2.41, 3.46)	0.977	0.14 (0.08, 0.26)	0.923
≥ 9	4.16 (3.15, 5.51)	0.491	0.39 (0.28, 0.54)	0.176
CRP level (mg/l)				
> 10	4.24 (1.16, 15.53)	< 0.001	0.11 (0.05, 0.25)	0.335

Values in parentheses are 95 per cent confidence intervals. LR, likelihood ratio; WBC, white blood cell count; PMN, polymorphonuclear; CRP, C-reactive protein; RLQ, right lower quadrant; LLQ, left lower quadrant. *Heterogeneity test.

Table 5 Discriminatory and predictive power of combinations of variables

	Reference	ROC area	Likelihood ratio		
			All variables absent	At least one variable present	All variables present
Guarding or rebound and WBC count $\geq 10.0 \times 10^9/l$	49	0.84 (0.80, 0.88)	0.14 (0.08, 0.24)	0.94 (0.72, 1.22)	11.34 (6.65, 19.56)
WBC $> 10.0 \times 10^9/l$ and CRP > 8 mg/l	57	0.96 (0.92, 1.00)	0.03 (0.00, 0.14)	0.53 (0.20, 1.37)	23.32 (6.87, 84.79)*
WBC $> 10.0 \times 10^9/l$ and CRP > 12 mg/l	32	0.85 (0.80, 0.90)	0.05 (0.01, 0.18)	1.07 (0.75, 1.47)	8.22 (4.73, 14.38)
WBC $> 12.0 \times 10^9/l$ and CRP > 6 mg/l	52	0.74 (0.66, 0.87)	0.06 (0.01, 0.33)	2.00 (1.45, 3.09)	—
WBC $> 10.0 \times 10^9/l$, CRP > 8 mg/l and IL-6 > 60 ng/l	57	0.87(0.80, 0.94)	0.03 (0.01, 0.16)*	2.06 (1.35, 3.25)	16.96 (3.08, 98.66)
WBC $> 9.0 \times 10^9/l$ and proportion of PMN cells $> 75\%$	55	0.66 (0.59, 0.73)	0.17 (0.07, 0.42)	1.54 (1.32, 1.79)	—
WBC $> 10 \times 10^9/l$, proportion of PMN cells $> 70\%$ and CRP > 12 mg/l	32	0.79 (0.74, 0.84)	0.03 (0.01, 0.16)	1.57 (1.28, 1.91)	20.85 (5.47, 80.27)
WBC $> 9.0 \times 10^9/l$, proportion of PMN cells $> 75\%$ and CRP > 6 mg/l	55	0.65 (0.58, 0.72)	0.05 (0.01, 0.28)*	1.44 (1.29, 1.63)	—
WBC $> 9.0 \times 10^9/l$, proportion of PMN cells $> 75\%$, manual bands $> 5\%$ and CRP > 6 mg/l	55	0.65 (0.58, 0.72)	0.05 (0.01, 0.29)*	1.43 (1.28, 1.62)	—

Values in parentheses are 95 per cent confidence intervals. ROC, receiver-operator characteristic; WBC, white blood cell count; CRP, C-reactive protein; IL, interleukin; PMN, polymorphonuclear. *One case was added to an empty cell for the calculations, giving underestimated ROC area and likelihood ratios.

Diagnostic value of combinations of variables

The diagnostic value of a combination of variables was analysed in nine studies. Andersson *et al.*⁴⁹ showed that a combination of inflammatory variables had the same discriminatory power for appendicitis as a combination of all the findings at clinical examination. Sex, signs of peritoneal irritation and inflammatory response variables were the independent predictors. Hallan *et al.*⁶² showed that inflammatory variables yielded additional diagnostic information over clinical and anamnestic variables. The ROC area of a logistic regression model increased from 0.85 to 0.92 when information about the WBC, proportion of PMN cells and CRP level was added to a model based on anamnestic and clinical variables. The independent predictors in a study by Dixon *et al.*⁵⁴ were tenderness in the right lower quadrant, abdominal rigidity, guarding, rebound tenderness, pain aggravated by coughing or movement, no previous surgery, pain of central onset, normal micturition, duration of pain and male sex. These authors did not include any inflammatory variable in their study.

Five studies analysed the results of combinations of two to four signs and inflammatory variables (Table 5)^{32,49,52,55,57}. The discriminatory and predictive power increased considerably when two or more variables

were combined into three diagnostic levels, representing an increase in the value of none, at least one, and all variables. Appendicitis was likely when two or more descriptors of inflammation were increased, with a LR⁺ of more than 10; it was unlikely when all markers of inflammation were normal, with a LR⁻ of less than 0.10.

Discussion

This review shows that each element of the history and of clinical and laboratory examinations is of weak discriminatory and predictive capacity. However, clinical diagnosis is a synthesis of information obtained from all these sources, and a high discriminatory and predictive power can be achieved by an accurate understanding of the relative importance of variables in combination. Contrary to common opinion, simple and easily performed laboratory tests of the inflammatory response appear to be at least as important as discriminators as the clinical descriptors of peritoneal irritation, especially in advanced appendicitis. When the values of two or more inflammatory variables are normal, appendicitis is unlikely. Conversely, appendicitis is very likely when the values of two or more inflammatory variables are increased. This result is consistent with the finding of a high proportion of

normal laboratory values in patients who were submitted to non-productive appendectomy^{71,72}. Interestingly, pain on rectal palpation has no discriminatory or predictive power. The opinion that rectal palpation is an indispensable examination in the diagnosis of appendicitis cannot be supported.

Many of the more important variables showed heterogeneous results between the studies; there are several possible reasons for this. Clinical assessment is a subjective appraisal of a patient's reaction to a surgeon's examination. This process cannot be standardized, which explains the low interobserver reliability of clinical findings⁷³. Another reason relates to the heterogeneous nature of study populations in respect of the proportions of patients with appendicitis and advanced appendicitis. This may reflect differences in indications for referral and for admitting patients with suspected appendicitis to hospital, or differences in definition of the appendicitis diagnosis. Finally, the probity of pooling heterogeneous results can be questioned and the estimated ROC areas and likelihood ratios may be considered biased, but this was at least partly taken account of by the wider confidence intervals of the random-effects model.

This review has demonstrated that elements of the disease history, clinical findings and results of laboratory tests are weak individual discriminators of appendicitis. However, in combination they provide high discriminating power. Laboratory tests of the inflammatory response, and the clinical descriptors of peritoneal irritation and migration of pain, are the strongest discriminators and should be included in the diagnostic assessment of patients with suspected appendicitis.

Acknowledgements

This study was supported by the Scientific Committee, Jönköping County, Sweden.

References

- Andersson RE, Hugander A, Thulin AJ. Diagnostic accuracy and perforation rate in appendicitis: association with age and sex of the patient and with appendectomy rate. *Eur J Surg* 1992; **158**: 37–41.
- Rasmussen OO, Hoffmann J. Assessment of the reliability of the symptoms and signs of acute appendicitis. *J R Coll Surg Edinb* 1991; **36**: 372–377.
- Wagner JM, McKinney WP, Carpenter JL. Does this patient have appendicitis? *JAMA* 1996; **276**: 1589–1594.
- Jones PF. Suspected acute appendicitis: trends in management over 30 years. *Br J Surg* 2001; **88**: 1570–1577.
- Hoffmann J, Rasmussen OO. Aids in the diagnosis of acute appendicitis. *Br J Surg* 1989; **76**: 774–779.
- Vermeulen B, Morabia A, Unger PF. Influence of white cell count on surgical decision making in patients with abdominal pain in the right lower quadrant. *Eur J Surg* 1995; **161**: 483–486.
- Feeley TM, McFarlane DE, Devlin HB. Acute appendicitis: are any investigations cost effective? *Ir Med J* 1982; **75**: 475–476.
- Williams R, Mackway-Jones K. Towards evidence based emergency medicine: best BETs from the Manchester Royal Infirmary. White cell count and diagnosing appendicitis in adults. *Emerg Med J* 2002; **19**: 429–430.
- Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978; **299**: 926–930.
- Hallan S, Asberg A. The accuracy of C-reactive protein in diagnosing acute appendicitis – a meta-analysis. *Scand J Clin Lab Invest* 1997; **57**: 373–380.
- Calvo Rigual F, Sendra Esteve S, Mialaret Lahiguera A, Montagud Beltran E, Llanes Domingo S, Medrano Gonzalez J. Valor de la proteina C-reactiva en el diagnostico de la apendicitis aguda en el nino. *An Esp Pediatr* 1998; **48**: 376–380.
- Chia YW, Carachi R, Armstrong AA, McGarry GW, Carrington D. Serum alpha interferon in children with right iliac fossa pain. *J R Soc Med* 1993; **86**: 259–260.
- Dickson AP, MacKinlay GA. Rectal examination and acute appendicitis. *Arch Dis Child* 1985; **60**: 666–667.
- Doraiswamy NV. Leucocyte counts in the diagnosis and prognosis of acute appendicitis in children. *Br J Surg* 1979; **66**: 782–784.
- Sanchez Echaniz J, Luis Garcia M, Vazquez Ronco MA, Mintegui Raso S, Benito Fernandez J, Lopez Alvarez-Buhilla P. Valor diagnostico de la proteina C reactiva en las sospechas de apendicitis aguda en la infancia. *An Esp Pediatr* 1998; **48**: 470–474.
- Jobst M. Zur Diagnostik der Appendizitis im Kleinkindesalter. *Zentralbl Chir* 1998; **123**(Suppl 4): 77–79.
- Jones PF. Acute abdominal pain in childhood, with special reference to cases not due to acute appendicitis. *BMJ* 1969; **i**: 284–286.
- Mollitt DL, Mitchum D, Tepas JJ III. Pediatric appendicitis: efficacy of laboratory and radiologic evaluation. *South Med J* 1988; **81**: 1477–1479.
- Peltola H, Ahlqvist J, Rapola J, Rasanen J, Louhimo I, Saarinen M *et al*. C-reactive protein compared with white blood cell count and erythrocyte sedimentation rate in the diagnosis of acute appendicitis in children. *Acta Chir Scand* 1986; **152**: 55–58.
- Rubin SZ, Martin DJ. Ultrasonography in the management of possible appendicitis in childhood. *J Pediatr Surg* 1990; **25**: 737–740.
- Samuel M. Pediatric appendicitis score. *J Pediatr Surg* 2002; **37**: 877–881.
- Wright JE. The polymorphonuclear leucocyte count as an aid to the diagnosis of acute appendicitis. *Med J Aust* 1969; **2**: 956–958.

- 23 Antonsen S, Qvist N. Evaluating blood leucocytes, blood neutrophils and plasma elastase in acute appendicitis. *Theor Surg* 1988; **3**: 17–22.
- 24 Fenyö G. Routine use of a scoring system for decision-making in suspected acute appendicitis in adults. *Acta Chir Scand* 1987; **153**: 545–551.
- 25 Goodwin AT, Swift RI, Bartlett MJ, Fernando BS, Chadwick SJ. Can serum interleukin-6 levels predict the outcome of patients with right iliac fossa pain? *Ann R Coll Surg Engl* 1997; **79**: 130–133.
- 26 Hallan S, Lange C, Asberg A. Abdominal skin temperature in acute appendicitis. *Br J Surg* 1995; **82**: 177.
- 27 Koslowski L, Schmolke M. Welchen Wert haben Leukozytenzahlung und Temperaturmessung bei der Diagnose der Appendizitis? Beobachtungen an 9561 Kranken. *Dtsch Med Wochenschr* 1969; **94**: 892–898.
- 28 Landau O, Watemberg S, Arber N, Subchi A, Lev D, Hutba S *et al.* Retrospective analysis of the benefit of various acute-phase reactants for the diagnosis of acute appendicitis. *Dig Surg* 1996; **13**: 457–459.
- 29 Lindberg G, Fenyö G. Algorithmic diagnosis of appendicitis using Bayes' theorem and logistic regression. In *Bayesian Statistics 3*, Bernardo JM, Degroot MH, Lindley DV, Smith AFM (eds). Clarendon Press: Oxford, 1988; 665–668.
- 30 Ring-Mrozik E, Naegele S, Soder J, Hecker WC. Der Wert der CRP – Untersuchung in der Differentialdiagnose der nicht-akuten Appendizitis. *Klin Padiatr* 1991; **203**: 377–380.
- 31 Schmidt KG, Klitgaard NA, Lund J. NBT-testen ved appendicitis acuta. *Ugesk Laeger* 1977; **139**: 139–144.
- 32 van Dieijen-Visser MP, Go PM, Brombacher PJ. The value of laboratory tests in patients suspected of acute appendicitis. *Eur J Clin Chem Clin Biochem* 1991; **29**: 749–752.
- 33 Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med* 1986; **15**: 557–564.
- 34 Amland PF, Skaane P, Ronningen H, Nordshus T, Solheim K. Ultrasonography and parameters of inflammation in acute appendicitis. A comparison with clinical findings. *Acta Chir Scand* 1989; **155**: 185–189.
- 35 Ingram RR, Mohammed R, Tillman J. C-reactive protein and acute appendicitis. *J R Coll Surg Edinb* 1988; **33**: 115–116.
- 36 Henneman PL, Marcus CS, Butler JA, Freedland ES, Wilson SE, Rothstein RJ. Appendicitis: evaluation by Tc-99m leukocyte scan. *Ann Emerg Med* 1988; **17**: 111–116.
- 37 Nase HW, Kovalcik PJ, Cross GH. The diagnosis of appendicitis. *Am Surg* 1980; **46**: 504–507.
- 38 Joffe SN, Louw JH. Squelching caecum in acute appendicitis. *BMJ* 1972; **ii**: 114.
- 39 Lane R, Grabham J. A useful sign for the diagnosis of peritoneal irritation in the right iliac fossa. *Ann R Coll Surg Engl* 1997; **79**: 128–129.
- 40 Andersson RE, Hugander A, Ravn H, Offenbartl K, Ghazi SH, Nystrom PO *et al.* Repeated clinical and laboratory examinations in patients with an equivocal diagnosis of appendicitis. *World J Surg* 2000; **24**: 479–485.
- 41 Pieper R, Kager L, Nasman P. Clinical significance of mucosal inflammation of the vermiform appendix. *Ann Surg* 1983; **197**: 368–374.
- 42 Andersson R, Hugander A, Thulin A, Nystrom PO, Olaison G. Indications for operation in suspected appendicitis and incidence of perforation. *BMJ* 1994; **308**: 107–110.
- 43 Barber MD, McLaren J, Rainey JB. Recurrent appendicitis. *Br J Surg* 1997; **84**: 110–112.
- 44 Cobben LP, de Van Otterloo AM, Puylaert JB. Spontaneously resolving appendicitis: frequency and natural history in 60 patients. *Radiology* 2000; **215**: 349–352.
- 45 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; **143**: 29–36.
- 46 Dujardin B, Van den Ende J, Van Gompel A, Unger JP, Van der Stuyft P. Likelihood ratios: a real improvement for clinical decision making? *Eur J Epidemiol* 1994; **10**: 29–36.
- 47 Albu E, Miller BM, Choi Y, Lakhapal S, Murthy RN, Gerst PH. Diagnostic value of C-reactive protein in acute appendicitis. *Dis Colon Rectum* 1994; **37**: 49–51.
- 48 Alshehri MY, Ibrahim A, Abuaisha N, Malatani T, Abu-Eshy S, Khairulla S *et al.* Value of rebound tenderness in acute appendicitis. *East Afr Med J* 1995; **72**: 504–506.
- 49 Andersson RE, Hugander AP, Ghazi SH, Ravn H, Offenbartl SK, Nystrom PO *et al.* Diagnostic value of disease history, clinical presentation, and inflammatory parameters of appendicitis. *World J Surg* 1999; **23**: 133–140.
- 50 Björkstén B, Wählby L. The nitroblue tetrazolium (NBT) test and white blood cell count in patients with acute abdominal pain, with special reference to acute appendicitis. *Acta Chir Scand* 1975; **141**: 65–68.
- 51 Bolton JP, Craven ER, Croft RJ, Menzies-Gow N. An assessment of the value of the white cell count in the management of suspected acute appendicitis. *Br J Surg* 1975; **62**: 906–908.
- 52 Davies AH, Bernau F, Salisbury A, Souter RG. C-reactive protein in right iliac fossa pain. *J R Coll Surg Edinb* 1991; **36**: 242–244.
- 53 Davies MS, Cunningham S, Cooke DA, Donaldson DR, Thomas MH. Is the hot appendix really hot? *Postgrad Med J* 1993; **69**: 862–864.
- 54 Dixon JM, Elton RA, Rainey JB, Macleod DA. Rectal examination in patients with pain in the right lower quadrant of the abdomen. *BMJ* 1991; **302**: 386–388.
- 55 Dueholm S, Bagi P, Bud M. Laboratory aid in the diagnosis of acute appendicitis. A blinded, prospective trial concerning diagnostic value of leukocyte count, neutrophil differential count, and C-reactive protein. *Dis Colon Rectum* 1989; **32**: 855–859.
- 56 Emery M, Jones J, Brown M. Clinical application of infrared thermography in the diagnosis of appendicitis. *Am J Emerg Med* 1994; **12**: 48–50.
- 57 Erkasap S, Ates E, Ustuner Z, Sahin A, Yilmaz S, Yasar B *et al.* Diagnostic value of interleukin-6 and C-reactive protein in acute appendicitis. *Swiss Surg* 2000; **6**: 169–172.

- 58 Fenyo G, Lindberg G, Blind P, Enochsson L, Oberg A. Diagnostic decision support in suspected acute appendicitis: validation of a simplified scoring system. *Eur J Surg* 1997; **163**: 831–838.
- 59 Gil VF, Peinado E, Obrador E, Reyes P, Perez Barba C, Merino J. Validez de las pruebas diagnosticas para confirmar o descartar una apendicitis aguda. *Med Clin (Barc)* 2000; **114**(Suppl 2): 48–51.
- 60 Gollidge J, Toms AP, Franklin IJ, Scriven MW, Galland RB. Assessment of peritonism in appendicitis. *Ann R Coll Surg Engl* 1996; **78**: 11–14.
- 61 Hallan S, Åsberg A, Edna TH. Estimating the probability of acute appendicitis using clinical criteria of a structured record sheet: the physician against the computer. *Eur J Surg* 1997; **163**: 427–432.
- 62 Hallan S, Åsberg A, Edna TH. Additional value of biochemical tests in suspected acute appendicitis. *Eur J Surg* 1997; **163**: 533–538.
- 63 Izbicki JR, Knoefel WT, Wilker DK, Mandelkow HK, Muller K, Siebeck M *et al.* Accurate diagnosis of acute appendicitis: a retrospective and prospective analysis of 686 patients. *Eur J Surg* 1992; **158**: 227–231.
- 64 Jahn H, Mathiesen FK, Neckelmann K, Hovendal CP, Bellstrøm Gottrup F. Comparison of clinical judgement and diagnostic ultrasonography in the diagnosis of acute appendicitis: experience with a score-aided diagnosis. *Eur J Surg* 1997; **163**: 433–443.
- 65 John H, Neff U, Keleman M. Appendicitis diagnosis today: clinical and ultrasonic deductions. *World J Surg* 1993; **17**: 243–249.
- 66 Middleton SB, Whitbread T, Morgans BT, Mason PF. Combination of skin temperature and a single white cell count does not improve diagnostic accuracy in acute appendicitis. *Br J Surg* 1996; **83**: 499.
- 67 Miskowiak J, Burcharth F. The white cell count in acute appendicitis. A prospective blind study. *Dan Med Bull* 1982; **29**: 210–211.
- 68 Raftery AT. The value of the leucocyte count in the diagnosis of acute appendicitis. *Br J Surg* 1976; **63**: 143–144.
- 69 Sondena K, Buan B, Soreide JA, Nysted A, Andersen E, Nesvik I *et al.* Rapid C-reactive protein (CRP) measurements in the diagnosis of acute appendicitis. *Scand J Clin Lab Invest* 1992; **52**: 585–589.
- 70 Thimsen DA, Tong GK, Gruenberg JC. Prospective evaluation of C-reactive protein in patients suspected to have acute appendicitis. *Am Surg* 1989; **55**: 466–468.
- 71 Gronroos JM, Gronroos P. Leucocyte count and C-reactive protein in the diagnosis of acute appendicitis. *Br J Surg* 1999; **86**: 501–504.
- 72 Gronroos JM, Gronroos P. A fertile-aged woman with right lower abdominal pain but unelevated leukocyte count and C-reactive protein. Acute appendicitis is very unlikely. *Langenbecks Arch Surg* 1999; **384**: 437–440.
- 73 Bjerregaard B, Brynitz S, Holst-Christensen J, Jess P, Kalaja E, Lund-Kristensen J *et al.* The reliability of medical history and physical examination in patients with acute abdominal pain. *Methods Inf Med* 1983; **22**: 15–18.