Papers

Randomised controlled trial of ultrasonography in diagnosis of acute appendicitis, incorporating the Alvarado score

Charles D Douglas, Neil E Macpherson, Patricia M Davidson, Jonathon S Gani

Abstract

Objectives To determine whether diagnosis by graded compression ultrasonography improves clinical outcomes for patients with suspected appendicitis.

Design A randomised controlled trial comparing clinical diagnosis (control) with a diagnostic protocol incorporating ultrasonography and the Alvarado score (intervention group).

Setting Single tertiary referral centre.

Participants 302 patients (age 5-82 years) referred to the surgical service with suspected appendicitis. 160 patients were randomised to the intervention group, of whom 129 underwent ultrasonography. Ultrasonography was omitted for patients with extreme Alvarado scores (1-3, 9, or 10) unless requested by the admitting surgical team.

Main outcome measures Time to operation, duration of hospital stay, and adverse outcomes, including non-therapeutic operations and delayed treatment in association with perforation.

Results Sensitivity and specificity of ultrasonography were measured at 94.7% and 88.9%, respectively. Patients in the intervention group who underwent therapeutic operation had a significantly shorter mean time to operation than patients in the control group (7.0 v 10.2 hours, P = 0.016). There were no differences between groups in mean duration of hospital stay (53.4 v 54.5 hours, P = 0.84), proportion of patients undergoing a non-therapeutic operation (9% v 11%, P = 0.59) or delayed treatment in association with perforation (3% v 1%, P = 0.45).

Conclusion Graded compression ultrasonography is an accurate procedure that leads to the prompt diagnosis and early treatment of many cases of appendicitis, although it does not prevent adverse outcomes or reduce length of hospital stay.

Introduction

Acute appendicitis is one of the commonest surgical emergencies. Simple appendicitis can progress to perforation, which is associated with a much higher morbidity and mortality, and surgeons have therefore been inclined to operate when the diagnosis is probable rather than wait until it is certain.[1] A clinical decision to operate leads to the removal of a normal appendix in 15% to 30% of cases (although the figure may be higher or lower in certain demographic groups).[1] This proportion may be reduced by observing equivocal cases for a period of time, a practice that seems to be safe for most patients.[2] Some cases of appendicitis may resolve spontaneously.[3 4] None the less, if a period of observation culminates in the diagnosis of a ruptured appendix, the patient may have suffered a poor outcome that was avoidable. Reductions in the number of "unnecessary" or non-therapeutic operations should not be achieved at the expense of an increase in number of perforations.[5]

It has been claimed that diagnostic aids can dramatically reduce the number of appendicectomies in patients without appendicitis, the number of perforations, and the time spent in hospital.[1] Methods advocated to assist in the diagnosis of appendicitis include laparoscopy,[6 7] scoring systems,[8 9] computer programs,[10] ultrasonography,[11] computed tomography,[12] and magnetic resonance imaging.[13] Imaging techniques have been shown to be particularly accurate.[14] Graded compression ultrasonography is the least expensive and least invasive of these and has been reported to have an accuracy of 71% to 95%,[14] but doubts have been raised about the influence of ultrasonography on patient outcomes.[15] Furthermore, it has been argued that findings at sonography should not supercede clinical judgment in patients with a high probability of appendicitis.[16] This raises questions about whether sonography should be performed at all in patients at high risk and whether there is some reliable means of selecting those who can benefit from imaging.

The Alvarado score is a 10 point scoring system for the diagnosis of appendicitis based on clinical signs and symptoms and a differential leucocyte count (see table 5). In his original paper Alvarado recommended an operation for all patients with a score of 7 or more and observation for patients with scores of 5 or 6.[8] Subsequent prospective studies have suggested that the Alvarado score alone is inadequate as a diagnostic test,[17 18] but it has been advocated as a means of selecting patients who should undergo imaging.[19]

We designed a diagnostic protocol incorporating graded compression ultrasonography and the

Department of Surgery, John Hunter Hospital, NSW 2310, Australia
Charles D Douglas
surgical registrar

Faculty of Medicine and Health Sciences, University of Newcastle, Callaghan NSW 2308, Australia
Neil E Macpherson
medical student
Patricia M Davidson
associate professor of paediatric surgery
Jonathon S Gani
senior lecturer

Correspondence to: CD Douglas cdouglas@ hunterlink.net.au

Alvarado score on the basis of work in our own institution.[20] We then undertook a randomised controlled trial to assess whether the information provided by the protocol improved clinical outcomes. We tested the hypotheses that compared with standard treatment patients assigned to the diagnostic protocol would have a shorter mean duration of hospital stay; a lower rate of unnecessary (non-therapeutic) operations; a shorter mean time to surgery for those undergoing therapeutic operations; and an equal or lower rate of delayed treatment in association with perforation.

Methods

Ethics committee approval was obtained for this trial. Patients were considered for inclusion in the study if they were referred to the surgical service at John Hunter Hospital and John Hunter Children's Hospital with a provisional diagnosis of acute appendicitis between October 1997 and October 1998. All general surgeons (six), paediatric surgeons (three), and their registrars (seven) at the participating hospital were involved in the study.

Patients were excluded from randomisation if they fulfilled any of the following criteria: age less than 5 years; evidence of generalised peritonitis; palpable mass in the right iliac fossa; evidence of acute confusional state or dementia; graded compression ultrasonography already performed. All other patients were met by the project officer, a third year medical student, who explained the study and obtained consent. The project officer randomly allocated patients by coin toss to control (standard treatment) or diagnostic protocol (intervention) groups. He organised a leucocyte count and performed a structured clinical assessment from which he calculated the Alvarado score (modified only by using percussion tenderness in the place of rebound tenderness).

For patients in the control group, members of the admitting surgical team were not informed of the Alvarado score. They proceeded with appropriate clinical assessment and management. They were requested not to organise graded compression ultrasonography for 36 hours.

For patients in the intervention group, the project officer advised the admitting team of the Alvarado score. Ultrasonography was then organised if the Alvarado score was between 4 and 8, inclusive. An Alvarado score of 9 or 10 was taken to be a relative indication for surgery, but the admitting team was given the option of organising graded compression ultrasonography; patients with an Alvarado of 3 or less were not eligible for ultrasonography. The admitting team was advised of the result of ultrasonography when this was done.

Ultrasonography

Graded compression ultrasonography results were designated positive, negative, or equivocal by the attending sonographer by using the following criteria: positive—appendix identified, tender and non-compressible or appendiceal phlegmon or abscess seen; negative—appendix not identified, no other relevant abnormality seen; equivocal—appendix not identified but abnormal amount of free fluid seen with thickened, dilated, or non-peristaltic bowel in the region of the caecum. In our experience these latter findings are often associated with perforation, and we suggested to participating surgeons and registrars that it was safest to consider such a report as a positive result.

In the few cases when the appendix was identified but was compressible or not tender we asked the sonographer to make a judgment on the basis of his or her experience and any other sonographic information, including appendiceal dimensions and blood flow. This reporting system was based on the results of a prospective study at our hospital (unpublished).

Ultrasonography was unavailable at this institution between the hours of 10 pm and 8 am, and patients entered in the study between these times had their examination at 8 am, unless the admitting surgical team deemed an immediate operation to be necessary.

Surgery

All patients who underwent laparotomy or laparoscopy for suspected appendicitis had an appendicectomy. The diagnosis of appendicitis was made on histological grounds on the basis of infiltration of the muscularis propria by neutrophil granulocytes.

A patient was considered to have had an operation if laparotomy or laparoscopy was performed. Operations were considered to be therapeutic if disease was found, when the disease seemed to be the cause for the patient's pain, and when surgery was the appropriate treatment for that disease. All other operations were classed as non-therapeutic operations. Operations were considered to be non-therapeutic if the only abnormality was a non-inflamed appendix containing a faecolith.

Perforations

The appendix (or bowel) was considered to be perforated if the surgeon clearly identified a perforation or a peritoneal swab grew at least one definite bowel organism or the histopathologist identified a perforation in association with gangrene or full-thickness necrosis.

Delayed treatment in association with perforation

For patients with perforation, treatment was considered to be delayed if surgery had not started within 10 hours of randomisation.

Follow up

Patients were reviewed at one week and three months with a pro-forma assessment. When direct review was not possible, details were obtained from the patient's general practitioner or surgeon.

Outcomes

We identified four outcome measures.

Time to operation for therapeutic operations was defined as the time in hours from randomisation to skin preparation.

Duration of stay was defined as the time in hours from randomisation to discharge from hospital. When a patient was discharged and then readmitted for ongoing management of a complication of acute appendicitis the duration of the subsequent admission was added to the first.

Rate of non-therapeutic operations—This was the number of non-therapeutic operations (see above) as a proportion of the total number in each group.

Rate of delayed treatment in association with perforation was the number of cases of delayed treatment in association with perforation (as defined above) divided by the total number in each group.

Power
This sample had a power of 80% to detect a difference between groups of 3.3 hours for mean time to theatre, 15.2 hours for mean duration of stay, and a reduction in the non-therapeutic operation rate from 11% to 2%.

Data analysis
Data were analysed on an intention to treat basis. For calculation of sensitivity and specificity of graded compression ultrasonography we included cases only if a histological diagnosis was available. Diagnoses other than appendicitis were ignored. Equivocal ultrasonography reports were counted as positive. Thus if the results of graded compression ultrasonography were reported as positive or equivocal for appendicitis but a perforated diverticulum and normal appendix were found at operation, the test was counted as a false positive, even though the operation was considered therapeutic.

Two by two contingency tables were analysed by Pearson's χ^2 test (or Fisher's exact test when stated), and comparisons of means were analysed by a two tailed t test. Confidence intervals for single proportions were calculated with the Wilson procedure without correction for continuity.[21]

Results

Figure 1 shows the trial profile. In total 306 patients were referred for inclusion; two patients failed to meet inclusion criteria and two patients refused consent, thus 302 patients were enrolled in the study, with 160 in the intervention group and 142 in the control group. The mean age was slightly lower in the intervention group (20.2 v 23.5 years); 202 patients were aged 14 years and over and 100 were aged under 14 years. There was little difference between groups with respect to sex, mean Alvarado score, or proportion with Alvarado score greater than 6. Figure 2 shows the distribution of Alvarado scores. Table 1 summarises the results. Subgroup analysis by age is shown in table 2.

Sixteen patients were in breach of the trial protocol because the admitting surgeon in each case thought that this was in the patient's interests. All were included in the reported analysis on an intention to treat basis. The results of a secondary analysis with these patients excluded were not substantially different with respect to adverse outcomes (intervention 17/154 (11%; 95% confidence interval 7% to 17%) v control 16/132 (12.1%; 8% to 19%); P = 0.8) or duration of stay (intervention 53.1 (46 to 60) hours v control 53.0 (44 to 62) hours; P = 0.99).

Ultrasonography
Graded compression ultrasonography was performed in 139 patients (see table 3). The sensitivity and specificity of ultrasonography for diagnosing appendicitis was 94.7% and 88.9%, respectively. There were three false negative results. Six patients with a positive or equivocal result on ultrasonography recovered without surgery.
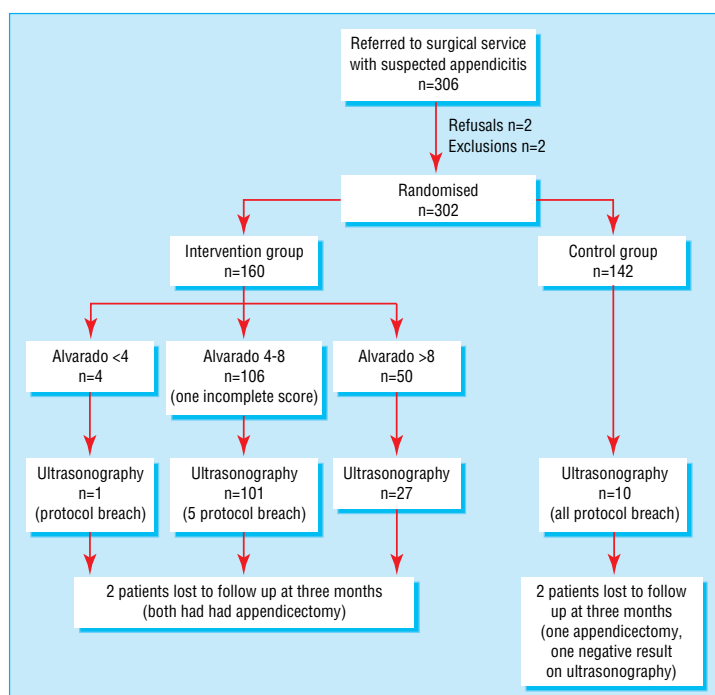
Fig 1 Profile of trial of graded compression ultrasonography and Alvarado score compared with clinical methods in diagnosis of appendicitis

Surgery
There were 170 operations performed: 95 of the 160 patients in the intervention group underwent surgery compared with 75 of the 142 patients in the control group (59.4% v 52.8%, P = 0.25). Appendicitis was confirmed histologically in 128 patients: 73 (45.6%; 38% to 53%) in the intervention group and 55 (38.7%; 31% to 47%) in the control group (P = 0.23). There were 13 patients with other conditions that met the criteria for a therapeutic operation (see table 4). Twenty nine operations were non-therapeutic: 14 (8.8%; 5.3% to 14.2%) in the intervention group and 15 (10.6%; 6.5% to 16.7%) in the control group (P = 0.59).

Perforations
Twenty four patients had perforations, 19 had a perforated appendicitis (14.8% (10% to 22%) of all cases of

Table 1 Demographics and results according to allocation to diagnosis with graded compression ultrasonography and Alvarado score or clinical diagnosis only. Results are given with 95% confidence intervals unless stated otherwise

	Intervention (n=160)	Control (n=142)
No male	87	69
Mean age (years)	20.2	23.5
Alvarado score		
Mean score	7.18	6.93
No (%) with score >6	99 (61.9)	86 (60.6)
Hospital data		
Total No (%) of operations	95 (59.4; 52 to 67)	75 (52.8; 45 to 61)
No (%) of therapeutic operations	81 (50.6; 43 to 58)	60 (42.3; 34 to 50)
Mean time to therapeutic operation (hours)	7.0 (5.9 to 8.1)	10.2 (7.9 to 13)
Mean duration of stay (hours)	53.4 (47 to 60)	54.5 (46 to 63)
No (%) of non-therapeutic operations	14 (8.8; 5.3 to 14)	15 (10.6; 6.5 to 17)
No (%) with delayed treatment in association with perforation	5 (3.1; 1.3 to 7.1)	2 (1.4; 0.4 to 5.0)
Total (%) adverse outcomes (delayed treatment in association with perforation and non-therapeutic operations)	19 (11.9; 7.7 to 18)	17 (12.0; 7.6 to 18)
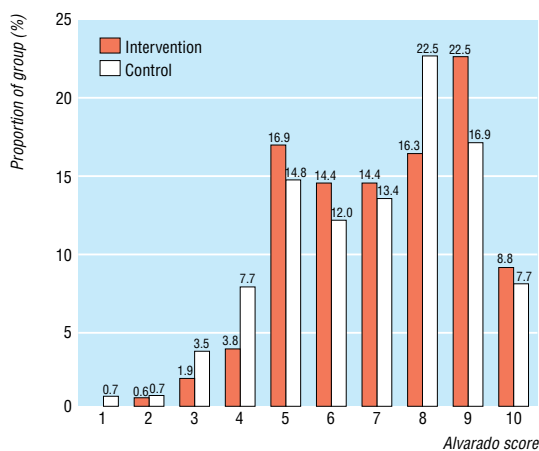
Fig 2 Distribution of Alvarado scores in patients with suspected appendicitis allocated to diagnosis with graded compression ultrasonography and Alvarado score (intervention group) or clinical diagnosis only (control)

appendicitis) and five had other bowel perforations. Of all perforations, 14 were in the intervention group and 10 in the control group (perforations/number in group of 8.8% (5.3% to 14%) and 7.0% (3.9% to 12%), respectively, $P = 0.58$).

Table 2 Primary outcomes in adults (>14 years) and children (<14 years) with suspected appendicitis according to allocation to diagnosis with graded compression ultrasonography and Alvarado score or clinical diagnosis only. Results are given with 95% confidence intervals

	Intervention	Control	P value
Adults (>14 years)			
No of patients	94	108	
Mean time to therapeutic operation (hours)	7.1 (5.8 to 8.4)	9.7 (7.2 to 12)	0.068
Mean duration of stay (hours)	59.3 (49 to 69)	59.2 (49 to 69)	0.99
No (%) of non-therapeutic operations	11 (11.7; 6.7 to 20)	12 (11.1; 6.5 to 18)	0.90
No (%) with delayed treatment in association with perforation	2 (2.1; 0.6 to 7.4)	1 (0.93; 0.2 to 5.1)	0.60*
Total (%) adverse outcomes (delayed treatment in association with perforation and non-therapeutic operations)	13 (13.8; 8.3 to 22)	13 (12.0; 7.2 to 20)	0.70
Children (<14 years)			
No of patients	66	34	
Mean time to therapeutic operation (hours)	7.0 (5.2 to 8.8)	13.1 (6.5 to 20)	0.069
Mean duration of stay (hours)	45.0 (37 to 54)	40.5 (27 to 55)	0.58
No (%) of non-therapeutic operations	3 (4.6; 1.6 to 13)	3 (8.8; 3.1 to 23)	0.41*
No (%) with delayed treatment in association with perforation	3 (4.6; 1.6 to 13)	1 (2.9; 0.5 to 15)	1.00*
Total (%) adverse outcomes (delayed treatment in association with perforation and non-therapeutic operations)	6 (9.1; 4.2 to 18)	4 (11.8; 4.7 to 27)	0.73*

*Fisher's exact test.

Table 3 Results of graded compression ultrasonography and sensitivity and specificity* for diagnosis of appendicitis

Result	Appendicitis (histologically confirmed)	Not appendicitis (histologically confirmed)	Not operated on	Total
Positive or equivocal	54	2†	6	62
Negative	3	16‡	58	77
Total	57	18	64	139

*Sensitivity 54/57 (94.7%; 95% CI 85.6% to 98.2%); specificity 16/18 (88.9%; 67.2% to 96.9%). Accuracy 70/75 (93.3%; 85.3% to 97.1%). Calculations based on histologically proved cases.
†Both equivocal reports: both had perforated colon.
‡Non-therapeutic operations.

Table 4 Pathological findings at surgery for therapeutic operations in patients with suspected appendicitis according to allocation to diagnosis with graded compression ultrasonography and Alvarado score or clinical diagnosis only

	Intervention group	Control group
Appendicitis	73 (46%)	55 (39%)
Diverticulitis	2	1
Tubal/ovarian abscess	2	
Perforated caecal carcinoma	1	
Ovarian cystic teratoma	1	
Perforated Meckel's diverticulum	1	
Omental infarction	1	
Acute cholecystitis		1
Mid-gut volvulus		1
Massive fimbrial cyst		1
Ectopic pregnancy		1

Delayed treatment in association with perforation

There were seven cases of delayed treatment in association with perforation (six cases of appendicitis and one of perforation of a caecal carcinoma). Five of these were in the intervention group and two were in the control group (3.1% v 1.4%, $P = 0.45$, Fisher's exact test).

Follow up

There were no readmissions with appendicitis during the follow up period. Two patients required readmission for complications: one in the intervention group for drainage of an abscess and one in the control group for an early small bowel obstruction. Nine patients had minor wound infections diagnosed one week after discharge (4 v 5, $P = 0.51$, Fisher's exact test). Four patients were lost to follow up at three months, three of whom had had their appendix removed during their admission (and were therefore able to be analysed for all end points); the fourth, in the control group, had had a negative result on ultrasonography before discharge.

Discussion

We have confirmed the high sensitivity and specificity of graded compression ultrasonography in the diagnosis of appendicitis. All our patients who underwent surgery after a positive result on ultrasonography proved to have appendicitis. Patients with equivocal signs of appendicitis are usually admitted to hospital for a day or night of observation. If the result on graded compression ultrasonography is positive, however, the surgeon can operate immediately. In our study, this lead to a significant reduction in mean time to therapeutic operation.

As some cases of appendicitis seem to resolve without surgery,[3 4] however, graded compression ultrasonography could lead to an increase in therapeutic operations (by correctly diagnosing appendicitis in patients who may have recovered during a period of observation). Our results are consistent with this, with a higher proportion of therapeutic operations occurring in the intervention group, although the difference was not significant. This is despite the fact that Alvarado scores were similarly distributed between groups.

The reduced time to operation in the intervention group did not result in a reduced duration of hospital

Table 5 Alvarado score for diagnosis of acute appendicitis

Symptom/sign/test	Score
Migration of pain	1
Anorexia	1
Nausea-vomiting	1
Tenderness in right iliac fossa	2
Rebound pain	1
Raised temperature (≥37.3°C)	1
Leucocyte count ≥10×10^9/l	2
Differential white cell count with neutrophils ≥75%	1
Total	10

stay. There was a trend towards shorter stays in the intervention group for patients undergoing therapeutic operations but because a larger proportion of the control group was managed non-operatively (and therefore discharged early) there was no difference overall.

Adverse outcomes

There are two outcomes that surgeons seek to avoid in cases of suspected appendicitis. The first is a non-therapeutic operation. The second is delayed treatment in a patient who is subsequently found to have perforation (delayed treatment in association with perforation). In this study, the proportion of patients in each group who had an adverse outcome (either a non-therapeutic operation or delayed treatment) was similar. The occurrence of a number of cases of delay with perforation, despite a low rate of perforated appendicitis (14.8%), suggests that rate of delayed treatment in association with perforation is a more appropriate measure of the consequences of delayed diagnosis than overall perforation rate.

In the intervention group, many adverse outcomes were associated with a clinical management decision that was at odds with a correct diagnosis under the intervention protocol. Of particular note is the fact that 11 patients had a non-therapeutic operation after a negative result on graded compression ultrasonography. If surgeons had relied on the results of the protocol these unnecessary operations would have been avoided. But what would have happened to other patients in the study?

The protocol, if allowed to override clinical judgment, would have led to operations in at least six patients who recovered without surgery and in two patients who did in fact undergo non-therapeutic operations (one with an Alvarado score of 10 and one with an incorrect sonographic diagnosis of ovarian cyst). More importantly, rigid adherence to the management indicated by the protocol would have required that no patient with a negative result could have an operation. There were three patients with appendicitis who had a negative result on ultrasonography, and in each case the appendix was gangrenous or perforated. Had graded compression ultrasonography been relied on, these patients would have had an indefinite delay in treatment.

Use of the Alvarado score to select patients for imaging

We used the Alvarado score as an objective means of stratifying patients according to risk so that those with a high or low probability of appendicitis need not have unnecessary imaging. The Alvarado score (table 5) is

based on a simple and largely objective assessment that requires minimal clinical experience, nevertheless NEM had an intensive period of training before the study began, and we ensured that he was eliciting symptoms and signs in a consistent and objective way. We believe that having one person perform the same objective assessment on all occasions was a reliable and reproducible method of risk stratification.

Whether the Alvarado score or some other form of risk stratification is used, selection of patients for imaging is an issue that cannot be ignored. Had we performed graded compression ultrasonography on all patients in the intervention group the results would probably have been worse. We would still have had three false negative tests, all in patients with a gangrenous or perforated appendicitis, and possibly more. Of the 31 patients in the intervention group who did not undergo graded compression ultrasonography there were 23 with Alvarado scores of 9 or 10, three with scores of 3 or less, and five breaches of protocol (because the surgeon was not prepared to wait for ultrasonography). Of these 31 patients, 27 underwent surgery and 25 (93%) had a therapeutic operation, with a mean time to operation of 5.6 hours. Our diagnostic protocol incorporating the Alvarado score was, if anything, safer, faster, and more accurate than graded compression ultrasonography alone, but it still failed to produce better outcomes than unaided clinical diagnosis.

Possible biases

Distributions of Alvarado scores in each group were similar, and the proportions of each group with an Alvarado score of greater than 6 (that is, patients who would be predicted to have appendicitis on the basis of Alvarado's original paper) were almost identical. Therefore the disparity between groups in number of

---

What is already known on this topic

Ultrasonography is an accurate test for the diagnosis of acute appendicitis

Few studies have examined the effect of diagnostic ultrasonography on clinical outcomes, and there have been no randomised controlled trials

What this study adds

This study confirmed the accuracy of ultrasonography and found a reduction in mean time to operation for patients undergoing therapeutic operation

There was no benefit of ultrasonography in terms of length of hospital stay, rate of non-therapeutic operations, or rate of delayed treatment in association with perforation

False negative tests occurred in patients with gangrenous and perforated appendixes

Ultrasonography remains a test of unproved benefit and should not be used by those who are inexperienced in the clinical diagnosis of appendicitis

therapeutic operations performed is unlikely to reflect a difference in disease prevalence.

To ensure that our results were not biased by the non-availability of the imaging service to after 10 pm, we performed a secondary analysis on the 231 patients who were enrolled in the study before this time. The difference in mean time to theatre was marginally short of significance and there was still no difference between groups in duration of hospital stay or in adverse outcomes.

General comments

When performed by experienced sonographers, graded compression ultrasonography is an accurate test. In this trial the accuracy was over 93%, equal to that of computed tomography without colonic contrast.[14] False negative reports, however, do occur: in our study 5% of negative results were incorrect. There is no certain way of determining which negative result is a false negative, and the consequences of not operating may be serious. Patients cannot be safely sent home after a negative result unless there are also clinical grounds for their discharge. It is therefore inappropriate for graded compression ultrasonography to be used by those who lack experience in the clinical diagnosis of appendicitis.

Conclusion

The diagnosis of acute appendicitis aided by graded compression ultrasonography has not been shown to produce better outcomes than clinical diagnosis alone. Further studies of graded compression ultrasonography and other diagnostic methods in suspected appendicitis should be randomised trials.

1   Hoffmann J, Rasmussen OO. Aids in the diagnosis of acute appendicitis. *Br J Surg* 1989;76:774-9.
2   Jones PF. Active observation in the management of acute abdominal pain in childhood. *BMJ* 1976;ii:551-3.
3   Ooms HW, Koumans RK, Ho-Kang-You PJ, Puylaert JB. Ultrasonography in the diagnosis of acute appendicitis. *Br J Surg* 1991;78:315-8.
4   Heller MB, Skolnick ML. Ultrasound documentation of spontaneously resolving appendicitis. *Am J Emerg Med* 1993;11:51-3.
5   Velanovich V, Savata R. Balancing the normal appendectomy rate with the perforated appendicitis rate: implications for quality assurance. *Am Surg* 1992;58:264-9.
6   Olsen JB, Myren CJ, Haahr PE. Randomized study of the value of laparoscopy before appendicectomy. *Br J Surg* 1993;80:822-923.
7   Moberg AC, Ahlberg G, Leijonmarck CE, Montgomery A, Reiertsen O, Rosseland AR, et al. Diagnostic laparoscopy in 1043 patients with suspected appendicitis. *Eur J Surg* 1998;164:833-40.
8   Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med* 1986;15:557-64.
9   Teicher I, Landa B, Cohen M, Cabnick LS, Wise L. Scoring system to aid in the diagnosis of appendicitis. *Ann Surg* 1983;198:753-9.
10 de Dombal FT, Leaper DJ, Staniland JR, McCann AP, Horrocks JC. Computer-aided diagnosis of acute abdominal pain. *BMJ* 1972;ii:9-13.
11 Puylaert JBCM. Acute appendicitis: US evaluation using graded compression. *Radiology* 1986;158:355-60.
12 Balthazar EJ. Appendicitis: prospective evaluation with high-resolution CT. *Radiology* 1991;180:21-4.
13 Incesu L, Coskun A, Selcuk MB, Akan H, Sozubir S, Bernay F. Acute appendicitis: MR imaging and sonographic correlation. *Am J Roentgenol* 1997;168:669-74.
14 Rao PM, Boland GWL. Imaging of acute right lower abdominal quadrant pain. *Clin Radiol* 1998;53:639-49.
15 Ford R, Passinault W, Morse M. Diagnostic ultrasound for suspected appendicitis: Does the added cost produce a better outcome? *Am Surg* 1994;60:895-8.
16 Sivit C. Imaging children with acute right lower quadrant pain. *Pediatr Clin North Am* 1997;44:575-89.
17 Ohmann C, Yang Q, Franke C. Diagnostic scores for acute appendicitis. *Eur J Surg* 1995;161:273-81.
18 Macklin CP, Radcliffe GS, Merei JM, Stringer MD. A prospective evaluation of the modified Alvarado score for acute appendicitis in children. *Ann R Coll Surg Engl* 1997;79:203-5.
19 Tobias M, Samuel E. Ultrasound and the Alvarado score. *J R Soc Med* 1992;85:585-6.
20 Davidson PM, Douglas CD, Hosking CS, Graded compression ultrasonography in the assessment of the "tough decision" acute abdomen in childhood. *Pediatr Surg Int* 1999;15:32-5.
21 Newcombe R. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med* 1998;17:857-72.

(Accepted 18 May 2000)