Annals of the Royal College of Surgeons of England (1990) *vol.* 72, 140–146

Twelve years' experience of computer-aided diagnosis in a district general hospital

William A F McAdam MB FRCS
Consultant Surgeon

Brenda M Brock BSc
Research Assistant for Computer-aided Diagnostic Project

Trevor Armitage MB FRCS
Surgical Registrar

Patricia Davenport
Medical Secretary

May Chan AFIMA
Project Officer, Clinical Information Science Unit

Francis T de Dombal MD FRCS
Director, Clinical Information Science Unit

Airedale District General Hospital, West Yorkshire

Key words: Acute abdomen, diagnosis, computers

This paper describes experience in a modern district general hospital with a small desktop system for computer-aided diagnosis of acute abdominal pain, over a 12-year period involving 5512 cases.

When compared with a baseline year (1973) in which unaided performance was monitored, during an initial study period (1974–76) the diagnostic accuracy of junior staff rose by between 10 and 15%. This higher performance level was then maintained for a decade (1976–86) despite changes in staff.

The perforation rate among appendicitis cases fell from 27% to 12.5%, accompanied by a smaller fall in negative laparotomy rates. The saving in surgical bednights devoted to acute abdominal pain was approximately 15%, and the notional cost of resources saved during the first 6 years of operation was £120 000.

Other hospitals have shown—in the short term—benefits similar to those obtained at Airedale District General Hospital. The long-term benefits of the system at Airedale reinforce the conclusions of the earlier short-term trials that a comparable system should probably be offered to all DGHs in the UK, not as an exercise in 'artificial intelligence' but as an effective continuing stimulus to good clinical practice.

Between 1966 and 1972, work in Leeds led to the development of a prototype system for computer-aided diagnosis of abdominal pain. This system was first evaluated by de Dombal *et al.* (*1*) and Horrocks *et al.* (*2*) in a series of consecutive patients presenting with acute abdominal pain to an academic surgical unit. In these 5512 patients the computer made less than half the diagnostic errors of the senior clinician who saw each patient. Subsequently, the same workers showed an improvement in diagnostic performance amongst junior clinicians who came into contact with a similar system (*3,4*).

These studies in Leeds left some important questions unanswered. Would such a system be transferable to a District General Hospital (DGH) environment? Would a 'less enthusiastic' environment—away from the originators of the system—nullify its effects upon the clinical performance? Would the placing of such a system in an 'ordinary DGH' environment have any tangible benefits? Finally, and perhaps most important in terms of wider implementation, would the effects 'wear off' with time?

In an attempt to answer these questions, the Leeds system was implemented in Airedale District General Hospital in 1974. The results of this implementation, and experience over a 12-year period, form the basis of the present report.

Aims and methods of study

Airedale DGH

Airedale District General Hospital is a modern DGH, operational since about the middle of 1970. It serves a

Correspondence to: W A F McAdam FRCS, Airedale District General Hospital, Steeton, Keighley, W Yorks, BD20 6TD

population of 162 140 persons in West Yorkshire (0.3% of the UK population). The staffing of the surgical unit consists of three teams, each of one consultant surgeon, one registrar or senior house officer and one preregistration house officer. Very few of the 77 persons involved in this study had previous experience of computer-aided diagnosis.

Patients studied

As in Leeds, *every patient presenting to Airedale DGH* (usually after direct referral by phone from their general practitioner, but occasionally directly via the accident and emergency department) with acute, undiagnosed abdominal pain of less than 1 week's duration, was entered into the present study, forming a consecutive, prospective, unselected series of patients from a well-defined population. A total of 5512 patients was thus studied.

Hardware/software

The system to be described used a desktop computer similar to that in Leeds, costing (in 1974) approximately £6000. This system comprised a small 8K desktop computer (a Wang 2200T), video screen and output writer for generating written case histories. Programs used were written in BASIC by members of the Leeds group. In February 1981 the original Wang was superseded by an Apple II costing £1500 and in 1987 by an Acorn Master costing around £900.

Years of study

The study reported here was divided into three time periods:

1 Baseline period (1973–74).
2 Study years (1974–76).
3 Service years (1976–86).

During the first period (baseline) performance of the junior hospital staff in respect of acute abdominal pain was monitored, but no assistance was provided. In the subsequent years (study years) the system was introduced, but was very much 'on trial'. In the first study year data was collected by the house officers, in the second by the SHOs. The former modus operandi—data collection by House Officers—was subsequently adopted. In both years the system was carefully monitored to ensure no harmful effects were associated with its introduction. Finally, in the subsequent 10 years the system as described was simply made available as a routine practice, and although some analyses were carried out, the intensive monitoring associated with the study years was not continued.

Running procedure

After admission to the ward, the patient was first seen by the house surgeon who recorded data on a specially designed form (Fig. 1). This form comprised the only written record about the patient's illness and was of the 'no-carbon-required' type. The top copy was retained in the case notes. The bottom copy was given to the computer operator, who processed the case as described earlier in the Leeds studies *(1)* and then attached the resultant printout with the computer's prediction to the patient's case notes.

Meantime, the SHO or registrar and/or consultant had seen and examined the patient, and a management decision had been made. No computer prediction was entered into the case records until after this immediate treatment decision had been recorded. The patient was then treated in a conventional manner and upon discharge the record was reviewed for such additional data as discharge diagnosis and length of stay. During the period 1975–77 a further review of non-surgical abdominal pain (NSAP) cases was carried out 1 year later through the patient's general practitioner.

As thus conducted, the Airedale experiment was designed to investigate the effects upon performance of DGH junior staff when a system was placed (without overt enthusiasm and without the involvement of the system originators in day-to-day running) in a DGH, over a period of several years.

Analysis

The method of analysis of these data closely followed that employed during the multicentre trial reported by Adams *et al.* *(5)*. The analysis concentrated upon the following features:

1 Technical performance of the computer systems.
2 Baseline performance data.
3 Effect of system on information collected.
4 Effect on diagnostic accuracy.
5 Effect on decision making and outcome.
6 Effect on resource utilisation.
7 Reaction on users.

Statistical analysis was carried out in the same way as in the multicentre trial *(5)* (see also *(6,7)*).

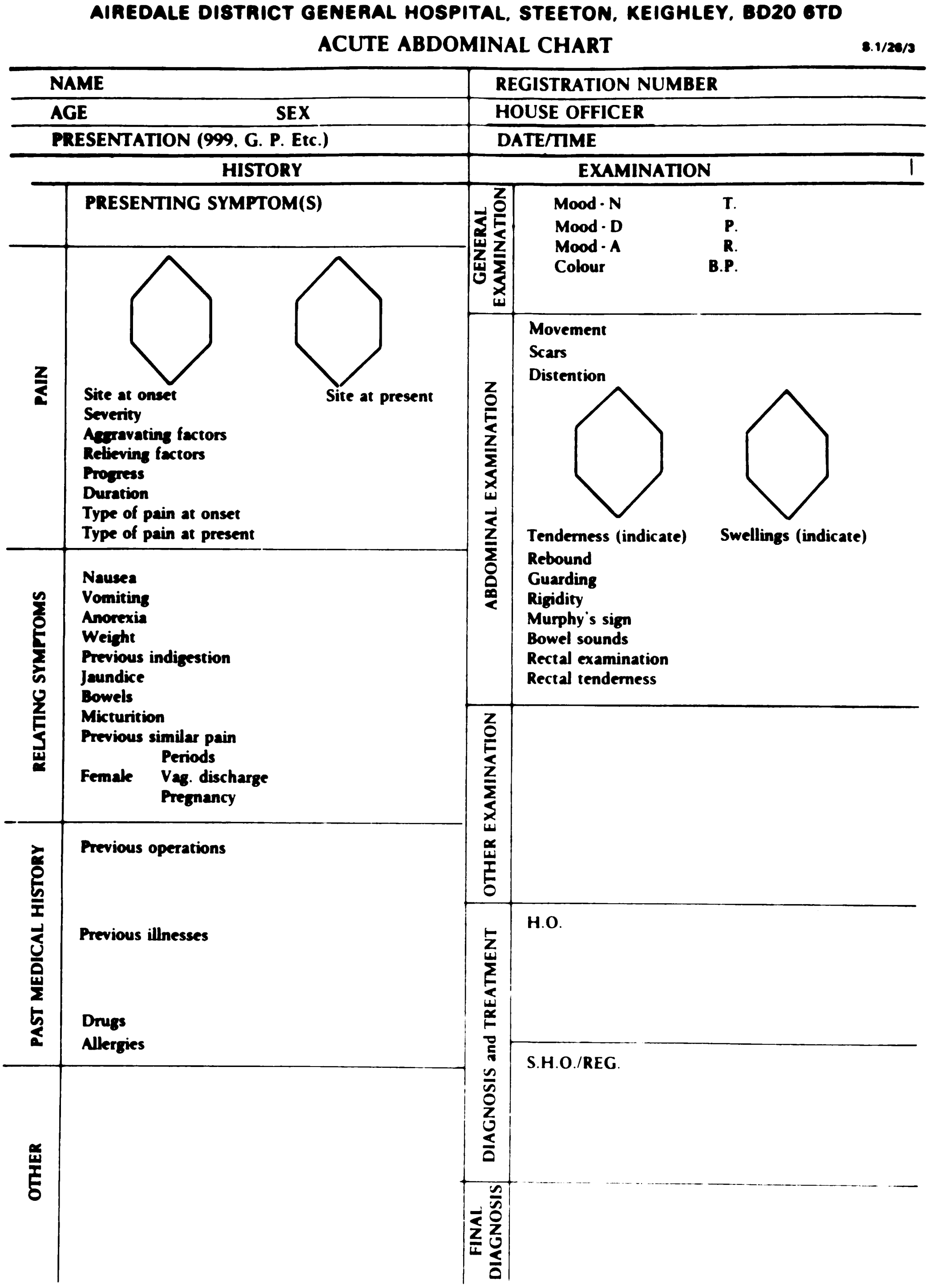
Results

Technical performance of system

The system has been available, installed on a variety of hardware (see above), for some 12 years (currently for 15 years). During that time, the longest interval for which it has been unavailable due to hardware or software faults has been 4 days. Overall, the system has been available for over 99% of the time.

Baseline performance data

Before installing the computer system, data was first collected concerning 397 patients presenting during 1973


AIREDALE DISTRICT GENERAL HOSPITAL, STEETON, KEIGHLEY, BD20 6TD

ACUTE ABDOMINAL CHART

8.1/28/3

NAME | REGISTRATION NUMBER

AGE SEX | HOUSE OFFICER

PRESENTATION (999, G. P. Etc.) | DATE/TIME

HISTORY | EXAMINATION

PRESENTING SYMPTOM(S)

PAIN
Site at onset Site at present
Severity
Aggravating factors
Relieving factors
Progress
Duration
Type of pain at onset
Type of pain at present

RELATING SYMPTOMS
Nausea
Vomiting
Anorexia
Weight
Previous indigestion
Jaundice
Bowels
Micturition
Previous similar pain
Periods
Female Vag. discharge
Pregnancy

PAST MEDICAL HISTORY
Previous operations
Previous illnesses
Drugs
Allergies

OTHER

GENERAL EXAMINATION
Mood - N T.
Mood - D P.
Mood - A R.
Colour B.P.

ABDOMINAL EXAMINATION
Movement
Scars
Distention
Tenderness (indicate) Swellings (indicate)
Rebound
Guarding
Rigidity
Murphy's sign
Bowel sounds
Rectal examination
Rectal tenderness

OTHER EXAMINATION

DIAGNOSIS and TREATMENT
H.O.
S.H.O./REG.

FINAL DIAGNOSIS

Figure 1. Data collection chart used for patients with acute abdominal pain, and forming part of the permanent case record during the period 1974–86.

and 1974 (Fig. 2). In these patients the diagnostic accuracy of the house surgeons was 54% and that of the senior house officers or registrars 58%. The perforation rate for patients with acute appendicitis (ie the proportion of patients with acute appendicitis whose appendix was found to have perforated at operation) was 27%, and the negative laparotomy rate, expressed as the proportion of patients with NSAP who were operated upon, was 22.3%. These figures are similar to those recorded originally in Leeds *(1)*, Scotland *(8)* and (more recently) elsewhere *(5)*.

Effect of system on information collected

Table I shows the effect of installing the system on the collection and recording for case notes of clinical data over the first 3 years of operation. Consideration is

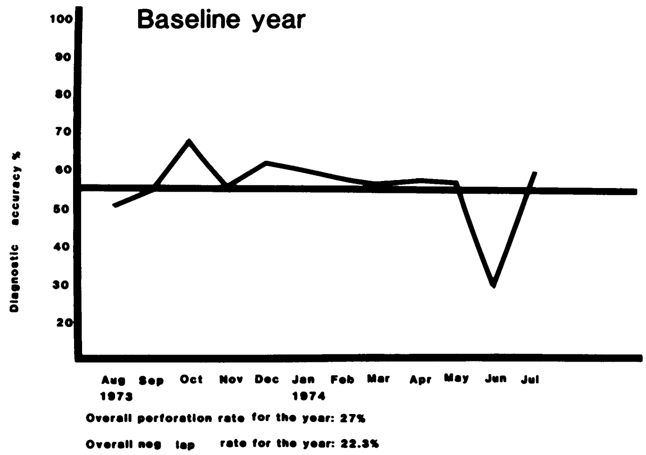


Figure 2. Diagnostic accuracy of house surgeon in first (baseline) year during which no assistance was provided. Note a complete lack of 'learning' during each of two 6-monthly firms.

restricted to attributes elsewhere shown to be crucial to the diagnosis in cases of suspected acute appendicitis *(9)*.

An immediate and substantial effect is seen upon the collection and recording of clinical data, the effect being maintained throughout the next decade.

This might be seen as directly attributable to the provision of data collection forms. However (Table II), it is clear that *data selection was also occurring*, with data items identified to the house surgeons and SHOs as crucial to the diagnosis of suspected appendicitis being missed only rarely. Other data, not unimportant but less crucial, were missed on up to 26% of occasions, even when data collection sheets were used.

Effect on diagnostic accuracy

As illustrated in Figs. 3 and 4, the accuracy of both house surgeons and registrars was maintained above baseline in each of the time periods under study, whether in the year of the initial experiment or in subsequent years when the system was in 'routine' service.

Overall, the house surgeons' accuracy rose from 54% in the baseline year to 66%; and that of the SHO/registrar from 58% in the baseline year to 71%. Both of these differences are statistically significant ($P<0.001$).

Table I. Effect of installing a system on clinical data collection/recording

	Time period (cases)		
	Baseline Feb–July '74 (240)	*Aug '74–July '75 (484)*	*Aug '75–July '76 (427)*
Site of pain	100*	100*	99*
Aggravating factors	46	90	92
Nausea	39	92	97
Vomiting	88	99	99
Appetite	59	93	94
Abdominal tenderness	100	99	100
Rebound	50	94	97
Guarding	63	96	99
Rectal examination	72	95	98
Amount of information on form collected and recorded	56%	91%	93%

* These figures indicate percentage of each group of patients where the particular item of data was collected and recorded by the first doctor who saw the patient

Effect upon decision making and outcome

Most workers in previous studies have chosen to evaluate three measures of decision making performance:

1 Perforation rates among appendicitis patients.

Table II. Showing 'selective' effect of impact on data collection. 'Vital' data to diagnosis of suspected appendicitis are omitted far less frequently than other less vital data

Missing data rate in 911 cases (August '74–July '76)			
*'Vital data' for diagnosis**		*'Less important' data†*	
Attribute	*% Cases NOT recorded*	*Attribute*	*% Cases NOT recorded*
Site of pain	0.4	Indigestion	16.7
Site of tenderness	0.6	Murphy's sign	18.1
Rectal examination	2.5	Distension	18.4
Guarding	2.6	Abdominal movement	19.1
Rebound tenderness	4.6	Subjective mood	26.6

* Items shown by previous studies *(1,9)* to have a high discriminatory value between appendicitis and non-surgical abdominal pain, listed as 'crucial' in the computer program provided *(3,4)*

† Items shown by some studies to have low discriminatory value between appendicitis and NSAP

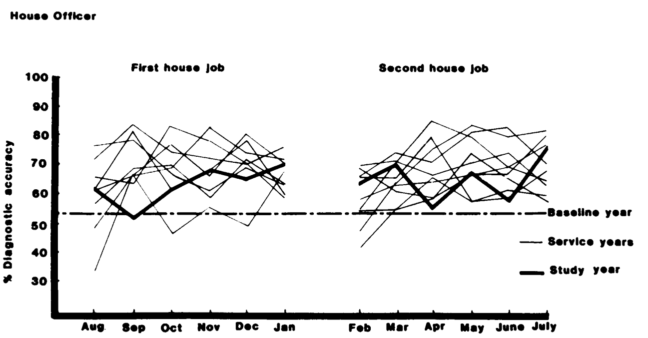


Figure 3. Diagnostic performance of house officers during subsequent years of trial. Note overall significant improvement compared with baseline year. (For definition of study year and service years, see text.)

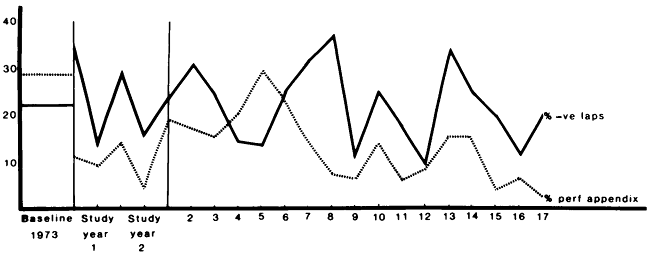


Figure 5. Negative laparotomy and perforated appendicitis rates during baseline year, study year and subsequent years. (Figures on horizontal axis indicate 6-monthly 'firms' during service years.) Note significant overall fall in perforated appendicitis rate, with smaller overall fall in negative laparotomy rate.

2 Negative laparotomy rates, and
3 Timing of surgery in patients with 'proven' acute appendicitis.

The results from the present study are illustrated in Fig. 5 and Table III.

With regard to perforation of appendicitis, in the baseline period the perforation rate (27%) was similar to that reported in other studies *(1,5,8,9)*. During the subsequent time periods, a total of 1005 appendicitis cases were observed, of which 126 (12.5%) perforated. This represents a substantial fall; significant at the 1% level. Moreover, as Fig. 5 illustrates, there has been no tendency to 'regress' with time. Indeed, the data from the last 2 years have been the best so far recorded.

In part, this significant drop in perforation rates may be due to the fact that in the test period patients with appendicitis have tended to proceed to surgery early. As shown in Table IV, 88.5% of patients came to surgery within the first 12 h of arriving in hospital; in part this may reflect provision of surgical facilities in a relatively new DGH; but it is noteworthy that the proportion of appendicitis patients whose operation was delayed for more than 24 h by diagnostic difficulty was 1.5%, as opposed to over 10% in the baseline period of the multicentre trial, and 12.2% in the Airedale baseline.

Figure 6 confirms this trend, analysing the prehospital and intrahospital delay among a group of 81 patients who perforated with acute appendicitis between 1978 and 1984, at which point this particular aspect of the study ceased. As shown, most of the delay in these patients took place outside the hospital. Indeed, only two patients with perforation were delayed over 24 h in hospital before operation out of 81 perforation cases, and a total of 498 patients with acute appendicitis surveyed.

Finally, negative laparotomy rates are also illustrated in Fig. 5. Here there has been a slight fall, though not as dramatic, from 22.3% in the baseline period to an overall value of 18.3% (359 out of the 1957 cases of NSAP who presented to hospital over the period of study). This fall is technically significant ($P<0.01$), but is far less marked than the fall in perforation rates for acute appendicitis patients. Nevertheless it does seem reasonable to conclude that the dramatic fall in perforation rates was not 'bought' at the expense of a concomitant rise in rates of negative laparotomy cases.

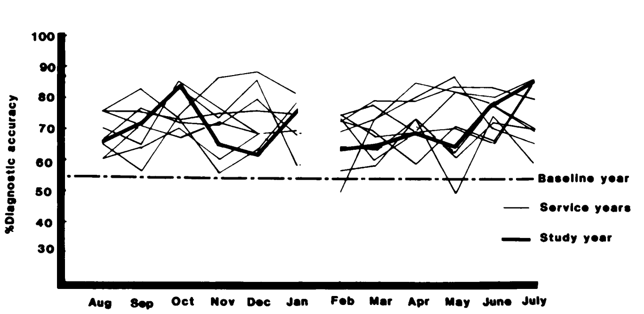


Figure 4. Similar analysis to *Figure 3* concerning SHOs and registrars. Note similar improvement overall.

Effect on resource utilisation

Resources in the National Health Service are notoriously difficult to analyse and quantitate. However, an attempt has been made to record simple parameters such as 'stay rates' for a variety of conditions as some measure of resource utilisation over the first 6 years of the study.

During that period (1974–80), stay rates for appendicitis patients fell from 8.2 to 6.8 days, while rates for NSAP patients fell similarly from 4.4 to 3.1 days. Using

Table III. Perforation of acute appendicitis during the 12-year period 1974–86

	Total appendix cases during period	*Cases perforated*	
		Expected	*Observed (%)*
Study period 1974–76	222	60	23 (10.4)
Next 5 years	360	97	68 (18.9)
Next 5 years	423	114	35 (8.3)
Overall (1974–86)	1005	271	126 (12.5)

* Based on both the 27% perforation rate among appendicitis patients in the Airedale baseline year (1973–74) and also the 27% perforation rate observed in the baseline period of the recent multicentre study of Adams *et al.* (5)

Table IV. Showing timing of surgery for acute appendicitis in Airedale during baseline and subsequent periods, with comparative data from multicentre study *(5)*.

	Multicentre study		*Airedale study*	
	Baseline (304 cases)	*Test period (1632 cases)*	*Baseline (105 cases)*	*Test Period (417 cases)*
1st 12 h	73.7%	82.8%	73.5%	88.5%
12–24 h	14.5%	11.9%	14.3%	10.1%
24+ h	11.8%	5.3%	12.2%	1.4%*

* Fall is statistically significant ($P<0.01$)

economic data provided and methods suggested for the multicentre trial *(5)* by the Office of Health Economics, the total value of the Average Total Recurrent Costs saved during the first 6 years of the trial (at 1981–82 prices) was around £120 000, and the Direct Cost saving around £25 000. These improvements appear to have been maintained up to date, though no formal assessment has been made.

This was not simply due to a policy change aimed at sending patients home earlier. The largest differences arose at the other end of the spectrum; for example the proportion of patients still in hospital 10 days after admission with complications of appendicitis and/or its treatment, fell from 25.4% in the baseline year to 11.6% in the subsequent 5 years ($P<0.01$).

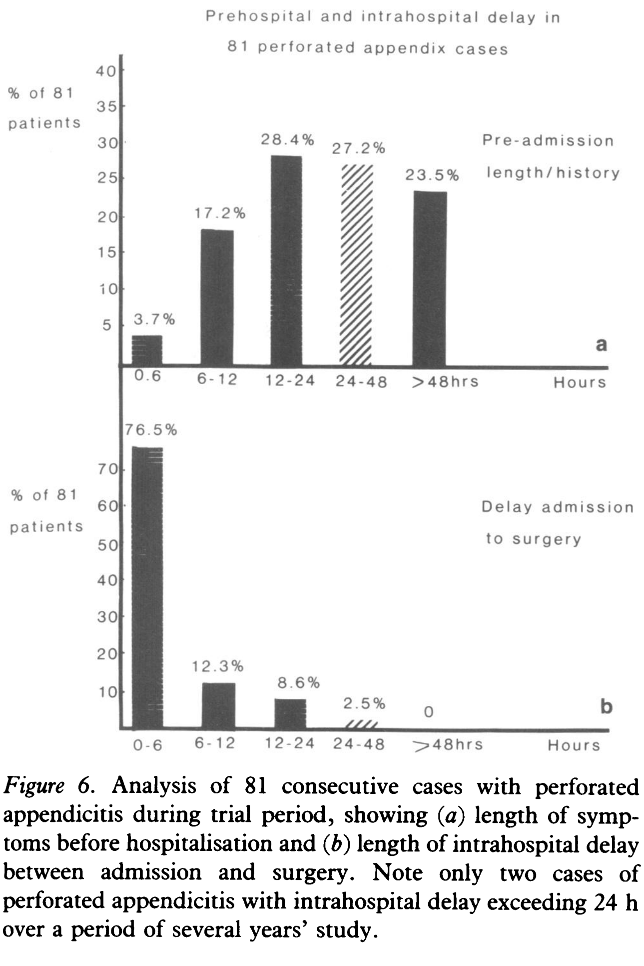


Figure 6. Analysis of 81 consecutive cases with perforated appendicitis during trial period, showing (*a*) length of symptoms before hospitalisation and (*b*) length of intrahospital delay between admission and surgery. Note only two cases of perforated appendicitis with intrahospital delay exceeding 24 h over a period of several years' study.

Discussion

Previous work in Leeds *(1)* has shown that whereas most junior hospital doctors have a diagnostic accuracy in the acute abdomen of between 45% and 55%, an accuracy of 70–80% can be achieved with care and the provision of a simple backup computer-aided system. This work has been amply confirmed by Gunn *(8)*, Wilson *(10)*, Boom *(11)*, Adams *et al.* *(5)*, Scarlett *et al.* *(12)* and Clifford *et al.* *(13)*. The present study conforms to this trend, and in this respect is of limited importance. However, two aspects of the present results are of considerable interest.

First, a conscious attempt was made to adopt 'a low profile' in installing this computer-aided diagnostic system. No effort was made to coerce the medical staff into using the computer themselves and no penalties were attached to the junior staff if they ignored the system totally. In view of this it is perhaps noteworthy that the system was adopted readily by the junior staff and, notwithstanding the 'lack of enthusiasm', the results of the system still appeared to be favourable.

Second, and perhaps more important, the results in year 12 were as good, if not better, in terms of accuracy and performance as the results in year one. This is important; the adoption of any novel device could lead to temporary enthusiasm and improvement. However, the Airedale studies have shown consistent improvement over the whole 12-year period and this finding has considerable importance for long-term use of such a system in the NHS. Indeed, the effects appear to have been maintained up to the time of writing (1989) though a full analysis has not yet been carried out.

It would be facile to pretend that the above improvements were due solely to the use of a computer, or even solely due to the direct feedback which the doctors received. As suggested elsewhere *(5)*, it seems reasonable to assume that at least half of the improvement was related to more thorough and disciplined data collection procedures.

This aspect was not directly addressed in the present study. However, some of the data do tend to suggest an effect over and above mere improvement in data gathering procedures. The evidence in Table II suggests that the doctor's mind was more focused on crucial clinical features than on others less crucial, and this suggests a 'selective' aspect to their data gathering which, had it occurred in the face of more conventional procedures, would have been called 'learning'. Moreover, when the computer system was briefly withdrawn in 1977 for a few months as an experiment, the doctors' diagnostic rate fell from 72.7% to 56.2% ($P<0.02$), the perforation rate for appendicitis cases rose from under 5% to over 20%, and the negative laparotomy rate rose from 9.1% to 22.6% ($P<0.05$).

The accuracy of the computer itself, though of only indirect relevance to the present study (no special effort being made to optimise this, for example by constant 'updating'), was approximately that of the improved performance by the SHO users. This experience parallels that of Boom *(11)*, Gunn *(8)* and Adams *et al.* *(5)*.

Airedale District General Hospital is, we believe, not untypical of many DGHs around the country, particularly those where newer buildings have been constructed in recent years. In this typical, modern DGH the improvements in clinical care, and decrease of undesirable events leading to unprofitable use of scarce resources have all been associated with the introduction of a simple computer-aided system.

There are considerable difficulties in assigning monetary values of 'savings' in the NHS and the probable value of the resources saved at Airedale must be notional. However, in the first 6 years' usage of the system, it would appear that this far outweighs the system running costs: a part-time operator at about £2000 per annum and amortisation costs of a few hundred pounds per annum. An improvement of 10% in clinical diagnostic and decision-making performance may not be very dramatic but it is nonetheless tangible, welcome and has been sustained over a 12-year period. If Airedale District General Hospital is typical of its counterparts, such systems, were they to be introduced more readily, might be expected to provide similar benefits, both clinical and in terms of usage of NHS resources.

In recent years there has been much talk of computer involvement in clinical medicine, nowhere more so than in the recent argument over increasing hospital 'efficiency'. In many instances doctors have been sceptical, wary of possible increasing dominance by some kind of 'artificial intelligence'. The Airedale system, dominated not so much by artificial intelligence as by the desire to 'do things right' clinically, probably represents a half-way house which is both acceptable and effective.

It is apparent that the work reported in this paper has been carried out by a large number of individuals and the authors extend to them their warmest thanks: To the senior colleagues for helpful input, to District Administration, to the Small Computer Advisory Group of the DHSS, to Skipton Rotary Club for funding and support, and particularly to the generations of young house officers and surgical staff whose experience forms the basis of this paper.

References

1 de Dombal FT, Leaper DJ, Staniland JR *et al.* Computer-aided diagnosis of acute abdominal pain. *Br Med J* 1972;2:9–13.

2 Horrocks JC, McCann AP, Staniland JR *et al.* Computer-aided diagnosis: description of an adaptable system and operational experience with 2034 cases. *Br Med J* 1972;2:5–9.

3 de Dombal FT, Leaper DJ, Horrocks JC *et al.* Human and computer-aided diagnosis of abdominal pain. Further report with emphasis on performance of clinicians. *Br Med J* 1974;**1**:376–80.

4 Wilson PE, Horrocks JC, Young CK *et al.* Simplified computer-aided diagnosis of acute abdominal pain. *Br Med J* 1975;2:73–5.

5 Adams ID, Chan M, Clifford PC *et al.* Computer-aided diagnosis of acute abdominal pain: a multi-centre study. *Br Med J* 1986;**293**:800–804.

6 Spiegelhalter DJ. Evaluation of clinical disease aids, with an application to a system for dyspepsia. *Stat Med* 1983; 2:207–16.

7 Pozen MW, d'Agostino RB, Selker HP *et al.* A predictive instrument to improve coronary care unit admission practices in acute ischemic heart disease. A prospective multicentre clinical trial. *N Eng J Med* 1984;**310**:1273–4.

8 Gunn AA. The diagnosis of acute abdominal pain with computer analysis. *J R Coll Surg Edinb* 1976;**21**:170–72.

9 de Dombal FT. The O.M.G.E. Acute Abdominal Pain Survey. Progress Report 1982. *Scand J Gastroenterol Suppl* 95 1984;**19**:28–40.

10 Wilson DH, Wilson PD, Walmsley GL *et al.* Diagnosis of acute abdominal pain in the Accident & Emergency Department. *Br J Surg* 1977;**64**:250–4.

11 Boom R. Improvement of internist's diagnostic performances by systematic computer evaluations. *Proc MEDINFO' 1980*, 1760.

12 Scarlett PT, Cooke WM, Clarke D *et al.* Computer-aided diagnosis of acute abdominal pain at Middlesbrough General Hospital. *Ann R Coll Surg Engl* 1986;**68**:179–81.

13 Clifford PC, Chan M, Hewett DJ. The acute abdomen: management with micro-computer aid. *Ann R Coll Surg Engl* 1986;**68**:182–4.

Received 19 October 1989