This paper describes a study of surgical diagnosis assisted by a computer. The study was carried out in the University of Leeds, involved the Department of Surgery and the Centre for Computer Studies and made use of an English Electric KDF9 computer.

A real-life real-time controlled trial was carried out on a consecutive unselected series of 552 further patients. This trial included all patients who presented to the Department of Surgery between January 1971 and August 1972 with abdominal pain of less than 1 week's duration. The pre-operative diagnosis of the most senior clinician who saw each case was accurate in some 81% of the patients studied. Using the same information (and producing its 'diagnosis' before the operation), the computer-aided system proved to be accurate in 91.5% of the cases.

The cost of each computer 'diagnosis' was around 10 new pence per case.

On these grounds it is suggested that the results of this initial study justify further modestly extended trials of the system.

**Introduction**

The idea of some form of automated system with the aid of which clinical diagnoses can be made is far from new. Indeed it has probably aroused more controversy than most other applications of digital computers in the biomedical sphere put together. Partly this controversy results from a misconception of the role of the computer in the diagnostic process. For, as Lusted (1966) cogently remarks, talk of 'computer diagnosis' or 'computers doing diagnosis' is not helpful at this time — and in Leeds the term 'computer-aided diagnosis' is preferred, implying that the central pivot on which medical decision making must rest is the diagnostic process of the clinician responsible for the management of the patient.

Can the computer help the clinician in this process? Here there have until recently been two main approaches, both complementary in their ultimate aims. The first, exemplified by the work of Professor Card (1967, 1970), inclines to the view that 'until we have formulated a theoretical structure, a logic, or a calculus of medicine ... it is not possible to transfer the activities of clinical medicine as we know them ... to a computer'; and this approach calls quite clearly for some detailed and fundamental research into the nature of the diagnostic process itself. Other workers (notably Edwards 1966), while accepting the need to formulate diagnosis in general terms, argue that in selected areas of clinical practice much can already be accomplished by utilizing available knowledge and techniques.

Our own view in Leeds is that not only are the two views complementary, but that recent events have led to a *rapprochement* of the twin pathways, so that current experiments in one area of activity may have a 'feedback' effect on both. Thus while our theoretical studies of the diagnostic process are detailed elsewhere at some length (de Dombal et al. 1971a, b, c, 1973; Leaper et al. 1972, 1973; Gill
et al. 1973) the present paper has quite a different purpose, namely the second of
the twin aims outlined above. That is to say, in the selected area of patients with
abdominal pain of acute onset, what can be achieved in routine practice using
presently available techniques and computing systems? This we attempted to assess
during an 18-month period, and our experience gained during this experiment
forms the basis of the present report.

**Clinical material**

We elected to study patients with abdominal pain of less than 1 week's duration
who presented to a busy surgical unit; and we chose this area of clinical medicine
for a number of reasons. First, such patients usually come to operation and thus a
detailed operative and histopathological diagnosis is usually available. Secondly,

<table>
<thead>
<tr>
<th>Table 1. List of disease categories considered by the computer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Such diseases account for 96% of all admissions with acute abdominal pain to</td>
</tr>
<tr>
<td>Professorial Surgical Unit, Leeds.</td>
</tr>
<tr>
<td>acute appendicitis</td>
</tr>
<tr>
<td>diverticular disease</td>
</tr>
<tr>
<td>perforated peptic ulcer</td>
</tr>
<tr>
<td>non specific abdominal pain</td>
</tr>
</tbody>
</table>

the problem of an 'acute abdomen' is an important and frequent one in clinical
surgery. Thirdly, in such circumstances, the clinician usually relies upon his own
interview and examination of the patient for diagnosis and decision-making,
rather than upon special tests; and finally it seemed probable that patients in large
numbers would be available for study.

Although the list of potential causes of abdominal pain is almost endless, a
preliminary survey revealed that in practice around 95% of patients presenting to
our unit might be expected to suffer from one of seven disease categories (those
listed in table 1). We therefore chose in the main to attempt a discrimination
between these seven common conditions and chose to accept that any 'rare'
complaint would automatically be mis-diagnosed by the computer-aided system.
To this end, we initially established a 'database' of clinical information concerning
100 patients with each of the seven 'common' diseases, and entered this informa­
tion into the computer before the start of the prospective trial.

Most of the disease categories in table 1 are self-explanatory, and most diagnoses
were made at operation, but one category – that of non specific abdominal pain
(or n.s.a.p.) – demands further explanation. 'N.s.a.p.', as Shepherd (1972) has
remarked, is scarcely a diagnosis, yet it is a concept familiar to every surgeon
experienced in abdominal surgery. It includes patients whose pain settles down –
and who are sent home often without any diagnosis being made; it includes patients
who came to 'negative laparotomy' – surgical exploration of the abdomen at
which no apparent disease is found; and finally the term n.s.a.p. embraces
occasional minor conditions for which no surgical measures are indicated – such as mild urinary tract infections.

**Computer-aided system**

At the outset it was decided that the computer-aided system should be capable of use within a routine clinical setting. This implied certain constraints upon the computing system; it became necessary to devise a system which would work around the clock, which contained an element of ‘backup’ capability, which would produce a diagnosis in ‘real-time’, i.e. before the patient went for operation, and also a system which the clinician could use without prior training or delay in his management of the case.

**Hardware**

The primary system which was devised (see Horrocks *et al.* 1972) made use of an English Electric KDF9 computer situated within the Centre for Computer Studies of the University of Leeds, about 750 m from the Department of Surgery. This central computer was accessed via a Westrex ASR 33 teletype located within the Department of Surgery using a G.P.O. Modem and Private Telephone Line linking the two departments. The primary system was available during University working days between 9.00 a.m. and 12 midnight. At other times a backup system involving a desk top computer (a Mathatronics 848 Biostatistician) was used.

**Software and systems**

As each case came into the ward, relevant details of clinical information were obtained from the Surgical Registrar when he first saw the patient. These details were copied onto specially designed forms (described in detail elsewhere, de Dombal *et al.* 1973) whose twin aims were to secure a maximum of useful and reproducible information about each patient. The information about each patient was encoded and entered into the computer, which produced its diagnostic prediction by comparing each new case with the ‘database’ of information about 600 patients using a variant of Bayes Theorem. The program by which this was accomplished was written by two members of our team (Dr A. P. McCann and Miss Jane C. Horrocks) in Fortran, and occupied less than 8 K of store in the computer. The patient’s name was never entered into the system for security reasons – each patient being identified by a series of code letters.

A sample of printout is shown in figure 1. This shows the patient’s reference number, the current symptoms, the computer’s diagnostic prediction, an attempt by the computer to resolve what is in this case a diagnostic discrepancy between its own ‘diagnosis’ and that of the clinician, and finally a list of rare diseases which may be of interest in an unusual case.
Female
Age..............20 TO 29
Site Onset......RIGHT LOWER QUAD
Onset Pain......OVER 48 HRS AGO
Type At Ons......COLICKY
Type Now......INTERMITTENT
Severity......PAIN NOW MODERATE
Nausea......PRESENT
Vomiting......PRESENT
Appetite......DECREASED
Bowel............NORMAL-NO CHANGE
Micturition......NORMAL
Jaundice......NOT PRESENT
Prev. Pain......YES-SIMILAR
Prev. Surg......NO PREV-ABD.OPRN
Drugs...........NO TREATMENT
Mood.............ANXIOUS
Colour.........FLUSHED
Abd. Movt......NORMAL
Abd. Scar......ABSENT
Distension......ABSENT
Tenderness......LOWER HALF ABD
Rebound......ABSENT
Guarding......PRESENT
Rigidity......ABSENT
Abd. Mass......NOT FELT
Murphy's Sign......NEGATIVE
Bowel Sounds......NORMAL
Rectal Exam......NO ABNORMALITY

Possible Diagnoses
Append Divert Peru Nonsap Cholec SMDOBT Pancre
Probabilities Are
25.44 0.13 0.00 73.74 0.00 0.69 0.00

Clinicians Diagnosis
Primary -Append
Secondary-Nonsap

Computers Diagnosis
Primary -Nonsap 73.74
Secondary-Append 25.44

Append is Certainly Possible but Nonsap and Append are Both Possible on This Data.

In Fact on Given Data Nonsap is Most Likely Diagnosis to Discriminate Between Append and Nonsap

++Suggest Checking the Following....
Guarding....
Rebound....
Aggrav. Fact.

Additional Diseases Worth Considering....
Pain Mainly Lower Abdomen SO:
Constipation
Crohn's Disease
Ulc. Colitis
Ca. Colon
Volvulus
Tuberculosis
Salpingitis(F)
Ovarian Cyst(P)
Urinary Tract Disease

Figure 1. Sample of printout from computer-aided system.
RESULTS

This system has been in operation since early 1971, and in all some 4000 diagnoses have been attempted in a variety of clinical areas. In respect of abdominal pain the system was run in conjunction with the clinical team for about 18 months. During this time we noted: (a) the admitting officer's diagnosis, (b) the house surgeon's diagnosis, (c) the registrar's diagnosis, (d) the diagnosis of the most senior clinician to see each case, and finally (e) the 'diagnosis' of the computer-aided system. (This last 'diagnosis' incidentally was not made available to the clinicians until after the patient had either gone to the operating theatre or the pain had settled down.) In all we studied some 552 cases, and this experience is now discussed.

Figure 2. Comparison of overall human and computer-aided diagnostic accuracy.

Overall diagnostic accuracy

This is shown in figure 2. As will be seen, the diagnosis with which the patient was admitted was (more often than not) erroneous, indicating that a considerable diagnostic problem yet remains when patients with abdominal pain arrive in hospital. The house surgeons fared rather better, as did the registrars and senior clinicians (whose overall diagnostic accuracy, 81.3% is above the national average, and improved from 79% to 83% during the trial). The overall diagnostic accuracy of the computer-aided system was however some 91.5%; its errors were significantly fewer than those of the clinicians ($P < 0.001$) and were largely confined to patients with rare diseases.
Comparative evaluation

The most common problem in acute abdominal surgery is that of 'appendicitis vs. n.s.a.p.'; in other words, does the patient have an acute appendicitis or a less urgent condition which does not require surgical intervention? This is therefore the sphere of activity which we have chosen to evaluate in some detail, and the results seem to us (table 2) to show a handy margin of performance in favour of the computer-aided system.

Table 2. Comparison of human and computer-aided diagnostic performance in discriminating between appendicitis and n.s.a.p.

Ignoring occasional 'other' diagnoses made by clinicians.

<table>
<thead>
<tr>
<th></th>
<th>correct append.</th>
<th>correct n.s.a.p.</th>
<th>false positive</th>
<th>false negative</th>
<th>overall % correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>receiving room</td>
<td>129</td>
<td>43</td>
<td>170</td>
<td>2</td>
<td>50</td>
</tr>
<tr>
<td>senior clinician</td>
<td>127</td>
<td>231</td>
<td>42</td>
<td>12</td>
<td>87</td>
</tr>
<tr>
<td>computer</td>
<td>141</td>
<td>262</td>
<td>8</td>
<td>—</td>
<td>98</td>
</tr>
</tbody>
</table>

As table 2 shows the diagnostic accuracy of the admitting officer in discriminating between these two common conditions was 50%. Such a level could be achieved by tossing a coin, but manifestly this was not so, since the table shows that the admitting officers were erring on the side of caution by admitting no less than 170 patients with 'appendicitis' whose pain rapidly disappeared on admission to hospital.

The senior clinicians fared better in discriminating between appendicitis and n.s.a.p. (87%). But even so there were 42 patients who fell into a ‘false positive’ category; they underwent a ‘negative laparotomy’ at which a normal appendix was removed and nothing else surgical was done. More serious, perhaps, was the finding that 12 patients with acute appendicitis fell into a ‘false negative’ category and a significant delay (mean 24 h, range 8 to 48 h) elapsed before their operation.

The computer-aided system discriminated between the two conditions with an accuracy of 98%. Moreover, although in some eight cases the computer-aided system made a ‘false positive’ diagnosis, there were no ‘false negatives’—so that had the clinical team been guided by the predictions of the computer some 34 patients would have been spared what in retrospect turned out to have been an unnecessary operation, and no patient with appendicitis would have experienced diagnostic delay before operation. This, as had already been remarked, constitutes what we regard as a handy margin of performance in favour of the computer-aided system.

Costs and evaluation

The cost of storing the relevant data and programs on the disk of the KDF9 and of running the program when required works out at around 10p for each new case run. To this must be added the time spent by the operator of the system in
the Department of Surgery, but since the data are elicited by the clinical team in the course of their routine duties, and since it takes less than 2 min to encode these data and enter them into the computer, these 'operating' costs are not great.

It is perhaps well that we should not evaluate the system ourselves, since we have an obvious interest in the matter. Yet we cannot resist pointing out two facts. First the cost of performing a 'negative laparotomy' and keeping the patient in hospital for a week postoperatively is rarely less than £200; and the system would have obviated the need for 34 such operations. More important perhaps, diagnostic delay which results in perforation of an appendix increases the mortality of appendicectomy tenfold; and the systems predictions would have helped to minimize such delay. The challenging assertion that diagnostic delay and 'negative laparotomies' are rendered unnecessary by the system is one we do not (yet) choose to make. But we feel it reasonable to assert that the running costs are minimal in view of the potential benefit of such a system.

Conclusions

In view of these data we have come to the following conclusions:

(1) We conclude that computer-aided diagnosis is potentially useful, is feasible, and is effective both in cost and diagnostic performance in respect of patients suffering from acute abdominal pain.

(2) It is necessary to enter the important proviso that these data refer to one clinical situation only, and it is still an open question whether comparable results can be obtained in different geographical areas and different spheres of clinical medicine.

(3) With this proviso, we conclude that further modestly extended trials of the system are probably justifiable.

The work was aided by a grant from the Medical Research Council which we acknowledge with gratitude.

References (de Dombal)

Edwards, W. 1966 Introduction to revision of opinions by men and machine systems. Institute of Electrical and Electronics Engineers (I.E.E.E.) Transactions on Human Factors in Electronics, Special Issue, pp. 1–6.


Shepherd, J. 1972 Correspondence, Br. med. J. 2, 347.