MEDICAL PRACTICE

Computers in Medicine

Clinical Diagnostic Process: An Analysis

D. J. LEAPER, P. W. GILL, J. R. STANILAND, JANE C. HORROCKS, F. T. DE DOMBAL

British Medical Journal, 1973, 3, 569–574

Summary

An analysis of observations made during 1,307 diagnoses by a total of 28 clinicians (503 diagnoses in real life, and 804 on simulated patients) concerned primarily the interview of patients suffering from abdominal pain. Interviews ranged from 10 to 35 questions, and from "stereotyped" procedures, in which identical (and often irrelevant) questions were asked to each patient, to "adaptive" interviews, in which specific relevant questions were put to each patient. Senior clinicians tended to ask fewer, more relevant questions than their junior counterparts; and urgent cases were dealt with in a more adaptive fashion than routine cases in outpatients. Disappointingly, there was considerable difference between real-life and simulated situations. From these results it is suggested (a) that the "diagnostic process" does not exist, (b) that any automated diagnostic system must be flexible to accommodate the wishes of a variety of clinicians, and (c) that studies based on artificial clinical situations should be treated with extreme caution.

Introduction

Several recent developments—such as the advent of the electronic digital computer and the widespread acceptance that "uncertainty" can be both measured and quantified—have revived interest in the "diagnostic process," and have led some to look forward to the emergence of a "technology" of clinical diagnosis.[1] But though there have been many studies of the ways in which the automated systems might make medical diagnoses, little is known about the way in which the clinician actually does so. We therefore set out in 1970 to study the clinician in rather more detail to assess the methods, techniques, and strategies which he uses in the clinical diagnostic process; and in this report we present in outline some of the findings from our survey.

Methods of Study

Any study which sets out to observe and analyse a real-life situation encounters serious problems, and we met with two which are worthy of special mention at the outset. Firstly, as soon as an observer is present—whatever the situation—he may alter (by his presence) the nature of the situation which he is trying to observe and document. Secondly, though a member of an experimental team may be present during the diagnostic process and "observe" what is going on, there is no guarantee whatsoever that his observations have any validity.[2]

INTRUSION OF OBSERVER

With regard to the first of these problems we initially evaluated several possible modes of observation using closed-circuit television, audiotape recorders, and so on. But the presence of such electronic equipment affected both the clinician and (equally importantly) the patient to a large degree, and we were unable to overcome this. For instance, after one particular interview recorded on audiotape both surgeon and patient inquired anxiously whether their "performances" had been "satisfactory"—a meaningless concept but one which illustrates the intrusive nature of this mode

University Department of Surgery, General Infirmary, Leeds LS1 3EX
D. J. LEAPER, M.B., CH.B., Research Assistant
P. W. GILL, M.B., CH.B., Research Assistant (At present Senior Registrar in Radiology, Meath Hospital, Dublin)
JANE C. HORROCKS, Computer Programmer and Physician's Assistant
F. T. DE DOMBAL, M.D., F.R.C.S., Reader in Clinical Information Science.
JOHN R. STANILAND, Research Assistant (Present address: Trinity Hall, Cambridge)

of observation. We therefore reverted to the use of a single human observer whose task was to work with the clinicians we studied, both on the wards and in the outpatient department, and to record in as discreet a fashion as possible what transpired during each diagnostic process. We hoped that in doing so any intrusion which might be felt initially by the clinicians would be short-lived, and indeed there is some evidence that this has been the case (fig. 1). Actually this arrangement worked better in most instances than might have been expected, partly because the clinician is accompanied by ancillary personnel (nursing staff, junior colleagues) more often than we had realized, and the presence of one more person rapidly became an unobtrusive one.

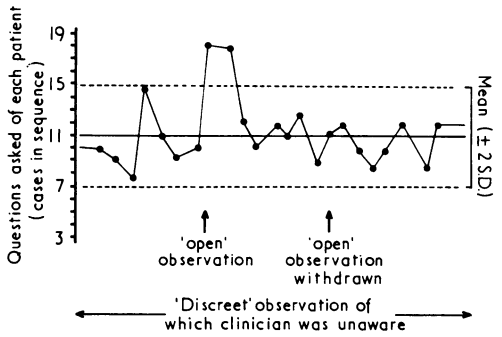

FIG. 1—Effect of observation on one aspect of clinician's diagnostic performance. Note short-lived effect of two additional obtrusive observers.

OBJECTIVITY OF OBSERVATIONS

We eventually decided to concentrate on the clinical interview rather than the physical examination since during the physical examination it is almost impossible to assess precisely what the physician is doing. (It should, however, be said that what analysis we have been able to make of the physical examination indicates that our observations "carry over" into this aspect of the diagnostic process also). We next decided to restrict our investigations to such observations as we felt able to make in an objective manner. This meant that we made no attempt to record or assess what Sprosty[3] referred to as "non-interrogative utterances" such as "O.K.," "all right," "uh," and "mmhmm" (sic). It also meant that we made no formal attempt at any psychological assessment of our colleagues; we merely graded them in terms of seniority and recorded what they did in objective terms. That is to say we took as an example a single clinical problem (patients with abdominal pain), and we recorded objective data such as what questions the clinician asked, the sequence in which he asked them, and the answer given by the patient.

We devised special forms for this purpose (fig. 2), and we carried out extensive observer variation trials to ensure that the data we recorded were reproducible.[4] Next, the actual recording of data was undertaken by only two of us (P.W.G. and D.J.L.), and we arranged for some clinicians to be "observed" by both observers. The eventual degree of variability between these independent observations was gratifyingly small.[4]

Finally, it should be stressed that while recording the data *we made no judgements* of the behaviour of our clinical colleagues. In the existing state of our knowledge of the diagnostic process, we felt that such judgements would have been both subjective and invidious, and we went out of our way to emphasize that we were merely anxious to document what was done rather than characterize it as "good" or "bad." This was much appreciated by our clinical colleagues.

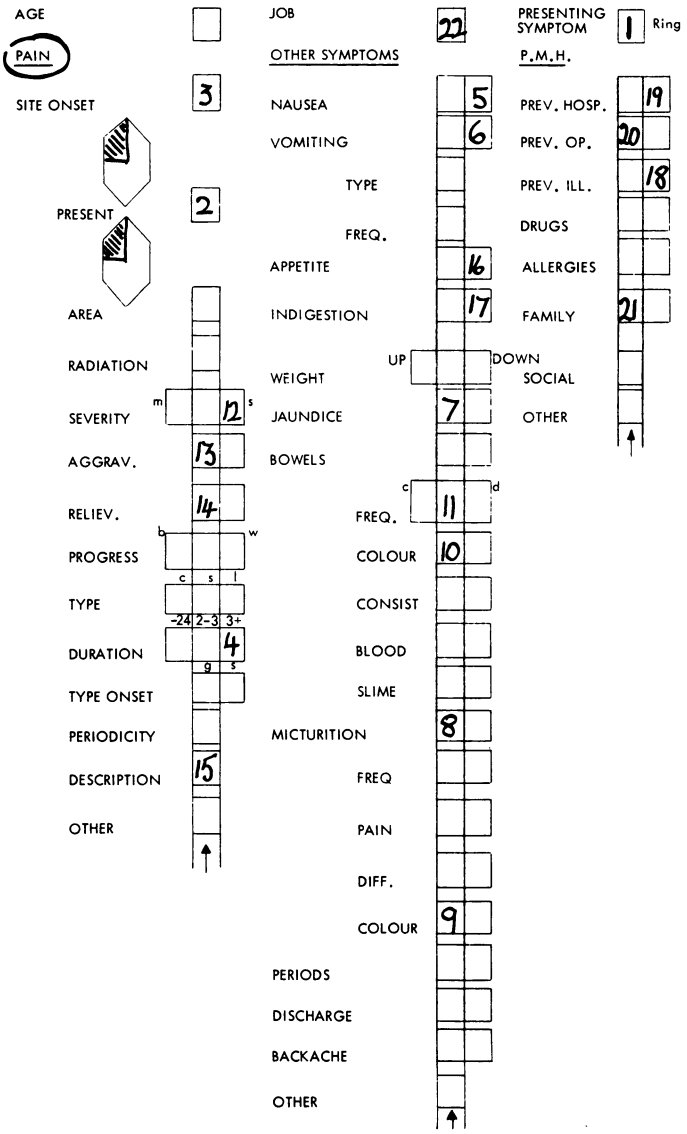

FIG. 2—Form devised for recording observations of diagnostic performance showing type of data recorded. Numbers indicate sequence in which questions were asked.

CONDUCT OF TRIAL

Between April 1970 and September 1972 28 clinicians participated in this study. Most were members of the Professorial Surgical Unit in the University of Leeds, but other surgeons at the Leeds General Infirmary, Leeds St. James (University) Hospital, and the Leeds Public Dispensary also participated. The breakdown by seniority of the surgeons taking part is shown in the table. In all some 1,307 diagnoses were observed, 503 in real life and 804 simulated diagnoses (the methodology of which has already been described elsewhere.[5 6]) Of the 503 "real life" diagnoses, some 221 were carried out on patients presenting with abdominal pain in the outpatient department, and the remaining 282 involved

Seniority of Surgeons Taking Part in Present Investigation

	No.	Percentage of Total
Consultant	4	14·3
Senior registrar	3 } 11	10·7 } 39·3
Registrar	8 }	28·6 }
House surgeon	13	46·4
Total	28	100

patients who were admitted as emergencies with acute abdominal pain. Each clinician was accompanied for two weeks before the recording of any data, in order to avoid the "intrusion" effect noted in fig. 1; and a "series" of observations on a single clinician comprised a maximum of 25 cases. The cases formed a prospective, unselected, and wherever possible a consecutive series, the data were recorded in "real-time," and no clinician's performance was discussed with him until after the closure of his "series."

Findings

In the analysis of this mass of data, we chose—for the reasons already mentioned—to make some extremely simple comparisons between the various groups of clinicians, and an outline of these overall results follows.

We shall in due course deal with the observations on simulated patients, and this initial phase concerns observations made solely in real life.

LENGTH OF INTERVIEW

Some variation from patient to patient is only to be expected, and fig. 3 shows the *mean* length of interview for a group of 28 clinicians. The mean length of interview in the group of clinicians we studied varied from around 10 questions to over 30. Nevertheless, quite evidently the length of interview is to some extent related to the experience and seniority of the clinician—in that the house officer tends to ask more questions than his senior colleagues. This is perhaps understandable, but what is surprising is the degree which we have observed in this trend. Equally interesting is the finding that there is little difference between registrar and consultant in this respect.

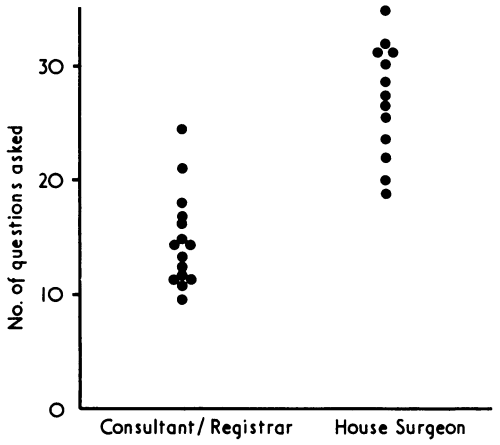

FIG. 3—Mean length of interview for a group of 28 clinicians. Each data point represents mean of 5 or more interviews for a particular clinician.

"TYPE" OF INTERVIEW

We made no attempt to assess in psychological terms the type of interview carried out by each clinician. We simply measured the degree to which each particular interview was "adapted" to the problems of each patient. That is to say, did the clinician ask the same questions in the same sequence on each occasion, irrespective of the precise diagnosis, or did he adapt his interview to the individual patient, asking "useful" and relevant questions? The results of this further analysis are shown in fig. 4.

Once again we found differences between house surgeons and their more senior colleagues. House surgeons tended to

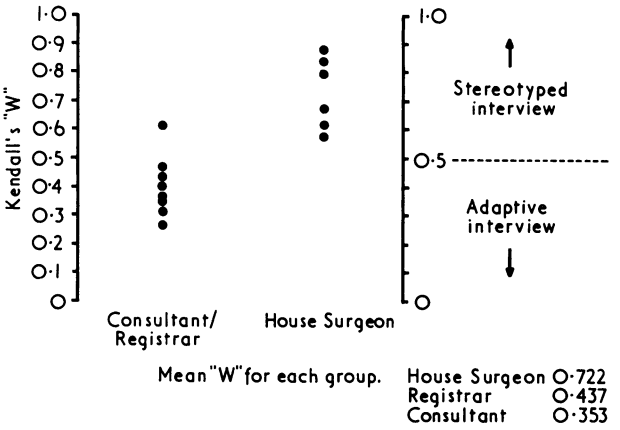

FIG. 4—Measurement of degree to which clinicians "adapt" their interview to specific patient situations. Each point represents Kendall's coefficient of concordance "W" for 20-25 interview "series" for a single clinician.

carry out a stereotyped interview, asking the same questions in the same order quite irrespective of whether they were useful or relevant to the particular case. The more senior clinicians on the other hand were much more likely to adopt an adaptive approach, interspersing general queries with specific questions designed to evaluate the presence or absence of particular complaint.

Not surprisingly, therefore, we found that the questions asked by registrars and by consultants had a higher "utility," particularly in the early part of the interview, than those asked by house surgeons (fig. 5). (There are actually several different ways of measuring "utility." We have illustrated in fig. 5 the analysis which results using an extremely simple concept based on the use of Bayes's theorem and maximum posterior probabilities, but it is worth pointing out that if one adopts a more complex mode of statistical analysis—such as "entropy" or "information energy"—then the findings and conclusions are strictly comparable.) In all, it is difficult to avoid the conclusion that the registrar and particularly the consultant are conducting a more "efficient" interview, gathering more useful information than the house surgeon, doing so in fewer questions, and having a higher overall diagostic accuracy.[7]

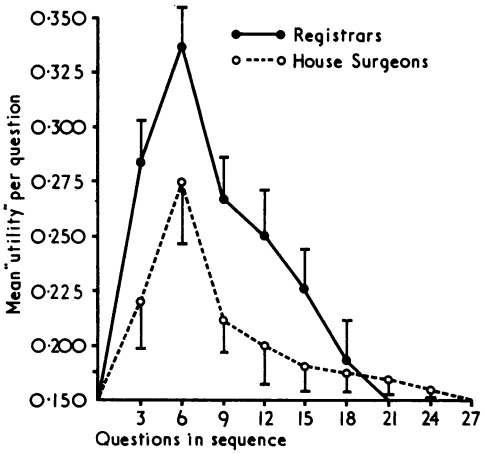

FIG. 5—Measurement of "utility" of questions asked by house surgeons and junior registrars.

EFFECT OF ENVIRONMENT

To the layman the hospital is a homogeneous environment, but manifestly this is not so, and the diagnostic process is indeed dependent in part on the environment in which it is carried out. The performances of clinicians in two situations,

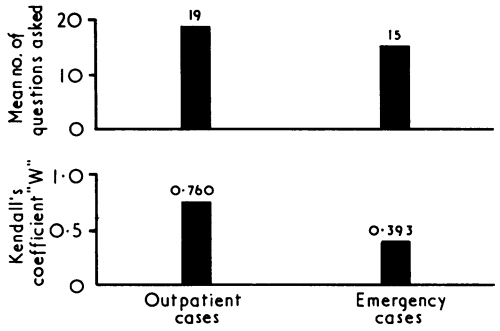

FIG. 6—Comparison of diagnostic performance of clinicians of registrar status in two differing clinical situations.

in the outpatient department and dealing with acute emergencies, are shown in fig. 6. We have chosen to analyse only clinicians of registrar grade partly to obtain an unbiased sample and partly since these are chiefly the clinicians who see patients in both clinical modes. Not surprisingly, in an emergency environment the registrar asks rather fewer questions. However, in an emergency environment the whole character of his interview becomes more "adaptive" and this difference will be commented on further below.

EFFECT OF SIMULATION

The analysis of the diagnostic process has thus far concentrated on what happens in real-life situations. We did, however, carry out several hundred simulated diagnoses, both computer-aided and non-computer-aided (fig. 7). These results can

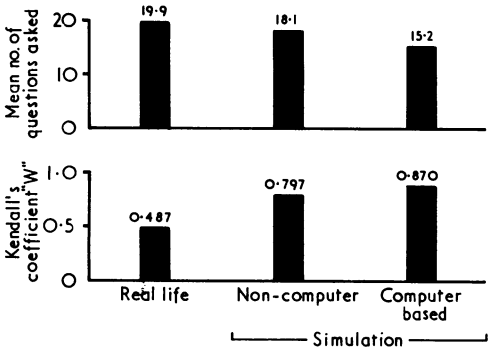

FIG. 7—Effect of simulation in varying mode on clinician, showing comparison with real-life performances.

only be described as thoroughly chastening since they confirm our earlier small-scale studies by showing that the clinicians' behaviour under simulated conditions was totally different from that in real life. Particularly noteworthy is the way in which clinicians who in real life conducted adaptive interviews, asking useful relevant questions, and who had a high diagnostic accuracy, showed none of these attributes when under simulated conditions. Without becoming involved in any semantic argument as to what constitutes a "good" or "bad" diagnostic process it is quite apparent that any results derived purely from artificial simulated diagnostic conditions must be viewed with extreme caution unless they are confirmed by parallel studies in real life.

Discussion

Of the various problems encountered and points raised during this study, perhaps three or four merit further discussion. Firstly, and by far the most important, is the implication that *the diagnostic process—viewed as a monolithic structure—does not in fact exist.* Each clinician has his own pathway to diagnosis, and while certain overall conclusions can be drawn, we certainly are in no position to draw up algorithms for different individuals which represent what they actually do. Not only does the diagnostic pathway vary from clinician to clinician, but from patient to patient—depending on such external factors as the "difficulty," urgency, and the role which the particular doctor assumes in the management of each particular case. Thus, for example, a clinician may behave quite differently in outpatients (where the patients are not acutely ill and where most of them are eventually referred for further investigations) and on the emergency ward (where many of the patients need operation within a few hours).

Such an observation explains the great difficulty encountered by our statistical colleagues in modelling and delineating in mathematical terms the diagnostic process—for in practice they are attempting to model something which does not exist as a single entity.

PHASES OF INTERVIEW

The shape of the registrars' profile in fig. 5 provides some confirmatory evidence for a model of the clinical diagnostic process which we put forward some two years ago.[8] Thus, for example, the most "useful" questions are not usually asked first (see fig. 5) but are asked part of the way into the interview. The effect of this can be seen in fig. 8, which analyses the performance of a registrar dealing with a patient suffering from acute appendicitis. The first few questions which the clinician asks do not greatly change the probability levels for the various common causes of acute abdominal pain. But at question 4 the clinician receives a strong indication that the patient may have acute appendicitis. Crucial to the present argument is the way in which this "clue" is turned to good account so that in the next few questions the diagnosis of appendicitis is confirmed beyond reasonable doubt. The remaining few questions are less concerned with the establishment of a diagnosis than with the subsequent treatment of the patient.

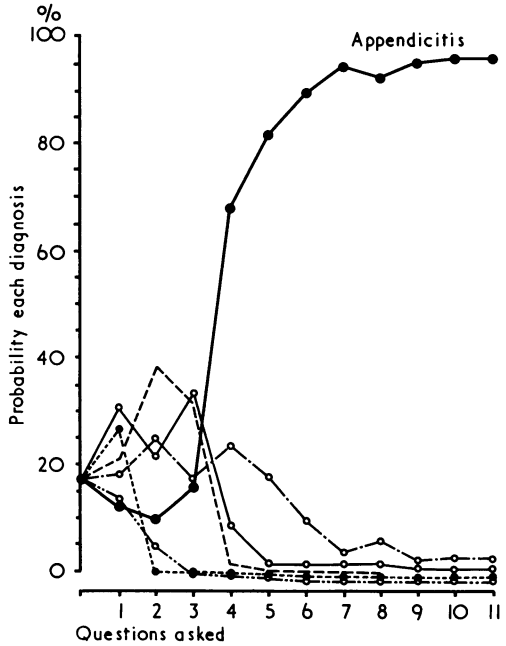

FIG. 8—Cumulative probability levels in an interview by registrar of patient suffering from acute appendicitis.

This procedure is representative of many interviews conducted by experienced clinicians faced with a reasonably straightforward diagnostic problem, and the structure of the interview is entirely in keeping with what we might have predicted after our earlier studies. Nonetheless, it is important to stress that not all interviews are conducted like this—particularly those conducted by more junior staff. An interview of the same patient carried out by a clinician with less experience is shown in fig. 9. As illustrated, at question one the clinician receives a comparable "clue" to the correct diagnosis—but owing to a more stereotyped approach he fails to follow up this clue with any further relevant questions. Indeed, as a result of the next few questions, the probability of acute appendicitis actually decreases and by the end of this particular interview the probabilities of a number of common diseases are virtually the same as those at the start of the interview. Not surprisingly this particular clinician failed to make the correct diagnosis in this instance. Thus, while our original model undoubtedly reflects accurately what happens in some diagnoses, it is a gross over-simplification of the clinical diagnostic process as a whole.

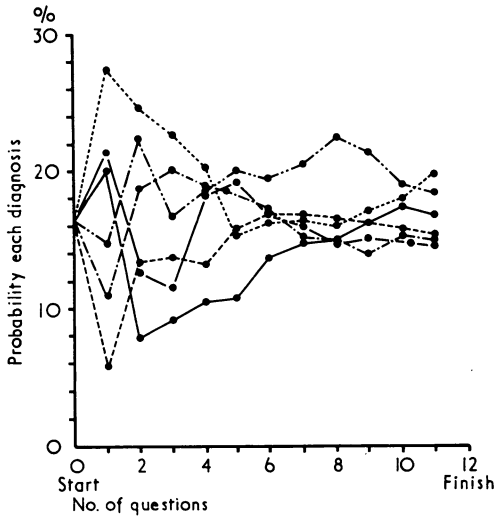

FIG 9—Same patient interviewed by house surgeon.

SIMULATION STUDIES *v.* REAL LIFE

We have written previously about the simulation of clinical diagnosis; and have entered some notes of caution regarding the validity of our studies. For, as Fraser[9] aptly remarks the question of fidelity of simulation is one which has plagued designers and users of simulation more than any other. What is both surprising and disappointing in the present series of observations is the degree of difference between behaviour observed under simulated conditions and behaviour observed in real life. Which is the "correct" version of a clinician's behaviour is perhaps irrelevant—far more to the point is the caution which such a finding places on any results emerging from purely artificial models of the diagnostic processes, including our own earlier studies[5 6]—though in our own defence it may be claimed that we did at least draw attention to the problem.

IMPLICATIONS FOR TEACHING

We said at the outset that we would make no judgements on our clinical colleagues, and while we do not in retrospect do so on individual clinicians, we feel that the results obtained in this study are such that some general comment must be made regarding the implications for teaching.

The house surgeon after five years' undergraduate training conducts a stereotyped interview which he is a priori totally incapable of assessing,[10 11] and often misses the significance of vital information purely because he is trained to obtain a case history in this stereotyped fashion. As we found in previous studies, "house surgeon" is a misnomer, since it appears to us that his role is restricted to "screening" the patient and consequently his behaviour is closer to that of a secretary than that of a surgeon[6]; and in retrospect we wonder if in fact five years' training is necessary to bring about such a state of affairs. Even more alarming was the unwillingness to embark on any sort of decision-making on the part of the house surgeons we studied. Indeed one of the clinicians whom we observed never made a decision during a series of over 20 cases.

The questions to be asked in respect of medical education are numerous. Do we wish to educate students to be highly-paid clerks in the first five to six years of their medical training? Moreover, in view of the fact that the senior clinicians' behaviour is not only more effective but also radically different from that of his junior colleagues, must we regard the latter behavioural state as necessary until the clinician has accrued some two or three years' postgraduate experience? Granted that it may be necessary to have some person "screen" the patient by asking 30 or 40 simple questions, is it really necessary to train a man for five years to do this or can this be accomplished by non-medically qualified "physician's assistants"? And, finally, if a clinician is to spend his life making decisions on which other people's lives depend, would it not be valuable to give him some training in decision-making, and give him some experience in making decisions at a relatively junior level?

IMPLICATIONS FOR DIAGNOSIS

The implications for diagnosis—particularly in respect of automated systems—are as follows. Firstly, it is quite clear that systems which adopt a rigid format will at best elicit a variable response from the clinicians who use them. Some clinicians may find that the system parallels their own variant of the diagnostic process. Others may find that this is so only part of the time and will on the whole tend to be less enthusiastic. Yet others may find that the system adopts an approach to diagnosis which is totally at variance with their own experience and practice—and it is extremely unlikely that such clinicians will welcome an automated system.

From the results found in this survey, any computer-aided diagnostic system which hopes to secure the active participation of a variety of clinicians must be *flexible*[7 12 13]—that is to say, such a system must be capable of accepting case histories of different length and in differing sequence; and it must be capable of providing an adaptable form of analysis of the case to suit the needs of an individual clinician.

We are grateful to Professor J. C. Goligher for his advice and encouragement throughout this survey as well as for permission to study patients admitted under his care. Three of us (D.J.L., P.W.G., and J.C.H.) were aided by a grant from the Medical Research Council for which we are also grateful. Finally, and particularly, we thank each of the 28 clinicians, who participated in this study and whose collective clinical experience forms the basis of our report.

References

[1] Lusted, L. B., in '*Introduction of Medical Decision Making*'. Springfield, Illinois, Thomas, 1968.
[2] Barsamian, W. E., in *The Diagnostic Process*, ed. J. A. Jaquez, p. 390. Ann Arbor, Malloy, 1964.
[3] Sprosty, P. J., in *The Diagnostic Process*, ed. J. A. Jaquez, p. 281. Ann Arbor, Malloy, 1964.

[4] Gill, P. W., *et al.*, *Methods of Information in Medicine*, 1973, 12, 108.
[5] de Dombal, F. T., Horrocks, Jane C., Staniland, J. R., and Gill, P. W., *British Medical Journal*, 1971, 2, 575.
[6] de Dombal, F. T., Horrocks, Jane, C. Staniland, J. R., and Gill, P. J., *British Medical Journal*, 1971, 2, 578.
[7] de Dombal, F. T., Leaper, D. J., Staniland, J. R., McCann, A. P., and Horrocks, Jane C., *British Medical Journal*, 1972, 2, 9.
[8] de Dombal, F. T., in *Principles and Practice of Medical Computing*, ed. L. G. Whitby and W. Lutz, p. 179. Edinburgh Churchill, Livingstone, 1971.
[9] Fraser, T. M., in *Philosophy of Simulation in a Man-Machine Space Mission System*. p. 35. Washington, N.A.S.A. SP-102, 1966.
[10] de Dombal, F. T., Horrocks, Jane C., Staniland, J. R., and Guillou, P. J. *Methods of Information in Medicine*, 1971, 11, 32.
[11] McCrae, A. W., *Psychological Bulletin*, 1970, 73, 112.
[12] Horrocks, Jane C., McCann, A. P., Staniland, J. R., Leaper, D. J., and de Dombal, F. T., *British Medical Journal*, 1972, 2, 5.
[13] de Dombal, F. T., Leaper, D. J., Horrocks, Jane C., Staniland, J. R., and McCann, A. P., *British Medical Journal*, 1973, In press.

Contemporary Themes

Psychogeriatric Ward for Mentally Disturbed Elderly Patients

D. M. PRINSLEY

British Medical Journal, 1973, 574-577

Summary

The operation is described of a special psychogeriatric ward of 23 beds set up in 1967 to provide treatment for mentally disturbed elderly patients who could not be kept in a general ward or at home. The unit is in a predominantly geriatric hospital which serves a population of 340,000 and in the four and a half years reviewed 600 patients were admitted. Half of the admissions were emergencies. A consultant geriatrician was in charge and the nursing staff were general trained. The number of beds was found to be adequate for the demand. Few patients had to be transferred to a psychiatric hospital, but, since the mental disturbance was often associated with severe illness and the patients were old, the death rate was high. The nursing staff have found the work interesting and stimulating.

Introduction

The incidence of organic brain disease in those aged over 65 has been stated to be 10%, and 40% of patients in geriatric wards may show severe psychiatric symptoms.[1] Because noisy patients in any ward disturb others they may be oversedated and thus under-diagnosed and under-treated. In 1967 a special psychogeriatric ward was opened at the Guisborough General Hospital, in the South Teesside Hospital Group, for the admission of elderly disturbed patients requiring any form of hospital treatment. Its effectiveness after five years is now clear, and this paper describes the unit and the patients it has dealt with in the period from its opening in May 1967 to the end of 1971.

The Psychogeriatric Ward

An unused, modern, well-constructed nurses' home at the hospital provided the sort of accommodation that was required—namely, many single rooms in a building relatively isolated from the main hospital but with access to full diagnostic and therapeutic facilities. Minimal alterations were made, call bells and observation windows were installed in each, sanitary facilities were improved, and a treatment room was provided. Much of the original furniture was retained, but conventional hospital beds were installed for the physically ill in rooms under direct observation from the sister's office. The ward was equipped to deal with acute physical illness in addition to disturbed behaviour. It consisted of 19 single and two double rooms, 23 beds in all, and three day/dining rooms. It was called Cleveland Ward.

There were no problems in accommodating both sexes and it was possible roughly to group the patients in the three day rooms according to their mental deterioration. There were no restraints but the hospital beds had slide-under cot sides. The front door was locked only when the unit housed a very active wanderer. A fenced garden with a circular path was later added for the benefit of patients who liked to wander. The treatment room was used for most diagnostic procedures and the patients' rooms remained free of medical equipment. The rooms adjoining the sister's office could be observed through large windows, and all patients were admitted to one of these for the first few days.

POPULATION SERVED

The South Teesside Hospital Group served a population of 340,000. The proportion aged over 65 was 10·6%, which was below the national average. Teesside family attitudes to elderly relatives were generally good; housing was conventional, with very few blocks of flats; there were many groups of sheltered housing with wardens; and the provision of homes for the elderly was about average. Industry was mainly heavy with no large-scale employment of women.

The average number of patients on general practitioners' lists was 2,632. The hospitals in the area catered for all specialties in 2,088 beds. The psychiatric hospital had 481 beds and there were 295 acute or long-stay geriatric beds.

Patients

In the period under review the unit treated 600 disturbed geriatric patients who were admitted because they could not be easily managed in an ordinary ward. Some of the patients

Department of Geriatric Medicine, Poole Hospital, Middlesbrough
D. M. PRINSLEY, M.D., F.R.C.P., Consultant Physician