MODELS OF THE DIAGNOSTIC PROCESS

by

THOMAS R. TAYLOR,
D.P.M.(London), M.R.C.P.(Glasgow and Edinburgh),
Lecturer in Medicine,
University Department of Medicine, Glasgow Royal Infirmary.

T H E S I S

submitted for the degree of

DOCTOR OF PHILOSOPHY

University of Glasgow

October, 1971.

C O N T E N T S

The studies presented in this thesis were carried out in the Department of Medicine, Royal Infirmary and Department of Medicine in Relation to Mathematics and Computing, University of Glasgow.

The work was performed when I was a Research Fellow supported by a Medical Research Council grant in the latter department and a Lecturer in the former department.

Data used in the studies was freely provided by Dr. J.A. Boyle, Department of Medicine, Royal Infirmary and Doctors W.D. Alexander and D. McLarty of Department of Medicine, Gardiner Institute, University of Glasgow.

The studies in Chapters 5 and 6 were performed in collaboration with Professor John Aitchison of the Department of Statistics, University of Glasgow. Programming assistance in all of the studies was provided by Miss Mary Nicolson, S.R.N. The studies in Chapter 7 were carried out in collaboration with Doctors W.D. Alexander and D. McLarty of the Gardiner Institute, University of Glasgow.

Some of the data has already been published in:

1. Taylor, T.R. (1970)
   J.Roy.Coll.Physn. Lond. 4: 188.

2. Taylor, T.R. (1970)
   Science Journal, 6, 81.

3. Taylor, T.R., Aitchison, J., McGirr, E.M. (1971)
   Brit.med.J. 3, 35.

PROLOGUE

The context in which the studies in this thesis are placed is that of decision-making in clinical medicine. Various aspects of clinical decision-making are examined and a number of tentative conclusions are advanced.

The first chapter paints in the background to decision-making in medicine and puts forward a possible mathematical framework for the investigation of the diagnostic process using information theory, conditional probability theory and decision theory.

In the second chapter a detailed critical historical review of studies in computer-assisted diagnosis is presented. The use of techniques of symbolic logic, probability theory, discriminant function analysis, decision theory and numerical taxonomy is illustrated.

The third chapter provides a simple description of the concepts of conditional probability theory, information theory and non-linear discriminant function which are used in the studies in this thesis; also included is a brief introduction to decision theory and to studies of clinical decision-making which have made use of its concepts.

There follows a group of studies of sequential decision-making using Bayes theorem and the entropy calculation of/

of information theory.

In the first of these studies in Chapter 4 the basic sequential model is described, in which the influence of prior probabilities, financial and non-financial costs, are discussed. Computer programs are presented, written by the author, which allow on-line computer diagnosis at a computer terminal. Significant reduction in the number of investigations needed to make a diagnosis in cases of non-toxic goitre is demonstrated.

Using this sequential probability model the studies covered in Chapter 5 provide a detailed diagnostic profile of each of six clinicians studying an identical set of cases of non-toxic goitre. This profile analysis is used to compare the clinicians indirectly with one another and to study the influence of personality factors and of clinical experience.

In Chapter 6 the diagnostic profile is presented on-line at two types of computer terminal. The relative merits of each terminal are compared and a system for the analysis and teaching of diagnostic skills is presented in some detail.

In Chapter 7 a quite different approach is made to the problem of selecting cases for treatment with anti-thyroid/

thyroid drugs. A non-linear discriminant function technique is used to distinguish the cluster of patients who are drug responsive from the cluster of those who are drug resistant. Only by taking into account the behaviour of such clusters over time was it possible to separate the two groups satisfactorily.

Finally, in the epilogue a number of possible avenues of further research are suggested which might produce answers to the many questions raised by the studies in the thesis.

THE DIAGNOSTIC PROCESS

C H A P T E R I : INTRODUCTION

I.1 The Diagnostic Process.

I.2 A mathematical analysis of the diagnostic process.

I.3 Sequential Decision-making.

I.1 The diagnostic process

Decision-making in clinical medicine has always been regarded as a distinctively human activity. However, the next decade will see digital computers increasingly used in the decision-making process that underlies diagnosis and treatment in clinical medicine. This promises to be perhaps the most exciting, and almost certainly the most significant, development in medicine today. It may well mark the end of an era of descriptive clinical medicine and the emergence of a new "twentieth century" objective clinical medicine. The assessment and management of patients as well as the decision-making which underlies these activities will change as the simple decisions are automated and insight is gained into the more complex ones.

A clinician faced with a diagnostic problem makes a sequence of decisions. He begins by choosing a symptom on which to concentrate initially and when he is satisfied about its presence or absence he chooses what he supposes will be the next most informative item. An experienced specialist repeatedly makes decisions about what to ignore in a patient's story. He decides when to ask the patient to expand his or her account of a symptom and where to concentrate his examination of the patient. In addition, he must decide what laboratory tests might be expected to be/

be most informative in a particular case and, finally, when he has gathered enough information he must decide what treatment to allocate and the probability of its success.

The selection and assessment of clinical information and the decisions about treatment are conducted against a background of the clinician's value system. Thus, he must decide if the expense and discomfort of any investigation or treatment is compensated by any expected improvement in the patient's health. Decisions of this type form the basis of the whole of clinical medicine. Yet the intellectual processes behind these decisions and even the premises on which they are based are poorly understood.

I.2 A Mathematical Analysis of the Diagnostic Process

Despite the enormous complexity of modern techniques of investigative medicine and of treatment, almost nothing is known of the intellectual processes by which investigations are selected, their results assessed and an appropriate treatment chosen. The diagnostic process is largely taught by example and almost no theoretical basis exists as yet to describe the intellectual activities which underlie it.

Techniques/

TRADITIONAL MODEL OF THE DIAGNOSTIC PROCESS

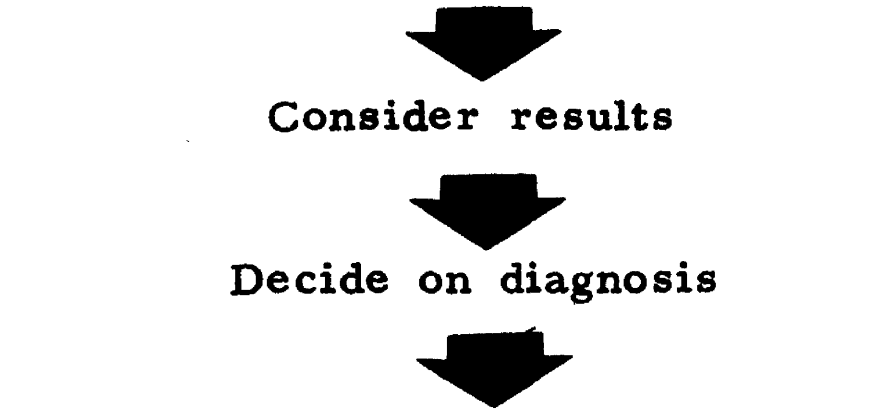


FIG. I.I. TRADITIONAL OR 'STATIC' MODEL OF THE DIAGNOSTIC PROCESS

Techniques of computer-assisted diagnoses have, until recently, been seen largely as interesting arithmetic exercises of little practical importance in medical practice. They have led, however, to the gradual emergence of mathematical concepts which describe important components of the diagnostic process (Taylor, 1967).

Almost all studies of computer-assisted diagnosis have attempted to reproduce the traditional or static view of the diagnostic process. The clinician is seen as collecting all available data on the state of health of his patient before assessing this, reaching a diagnosis and choosing a suitable treatment (Fig. 1.1). As clinicians we teach this version of the diagnostic process to our students but in practice we use a sequential form of decision-making (Ledley and Lusted, 1959; Ledley, 1966; Hamilton, 1966).

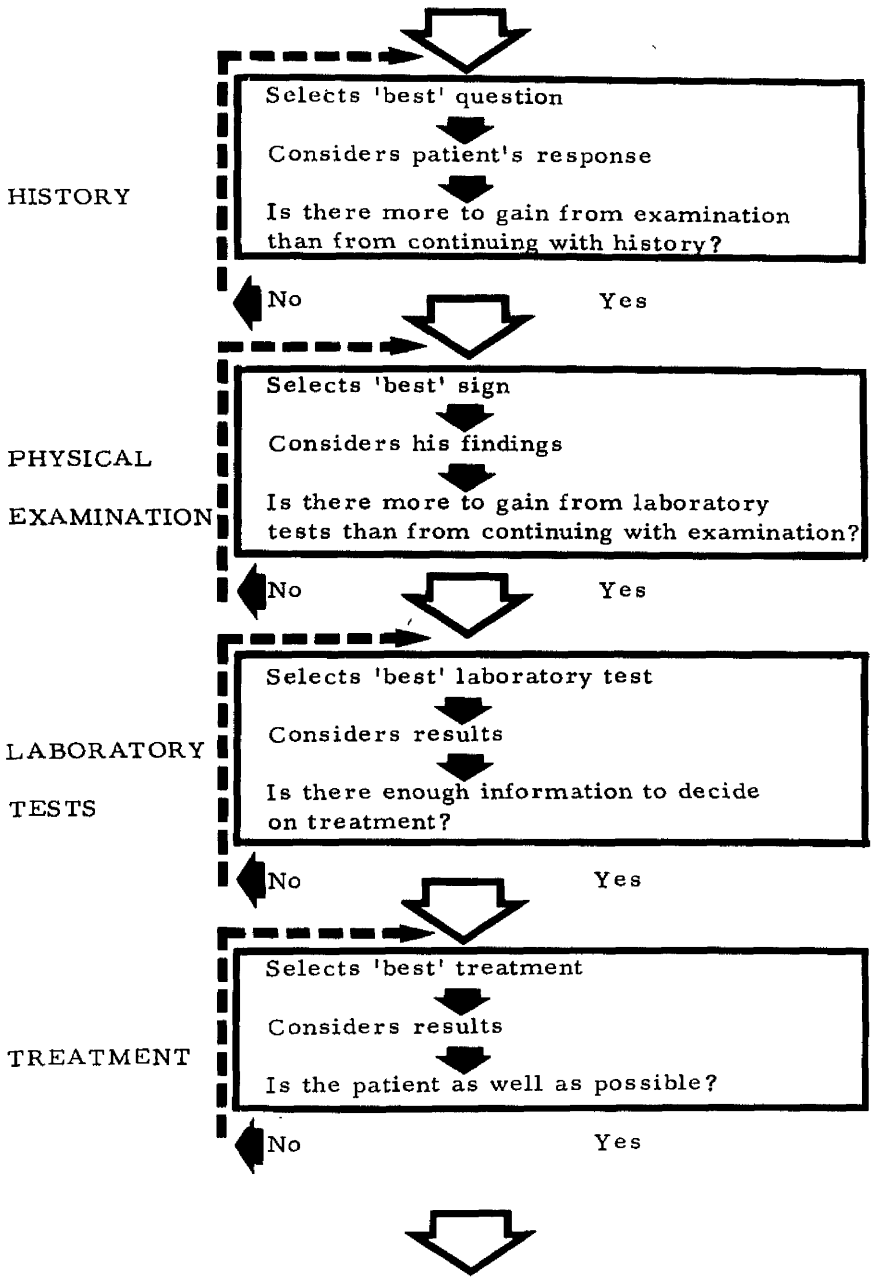


FIG. I.2. SEQUENTIAL DECISION MODEL

I.3 Sequential decision-making

After the diagnostic process has been broken down into a sequence of decisions (Fig. 1.2) it may be restated in mathematical terms by making use of information theory, conditional probability theory and statistical decision theory. Although clinicians are, of course, familiar with the use of such measurements as blood pressure, pulse rate and blood sugar level, they are largely unaware that further statistical techniques can be applied to the objective interpretation of such measurements. Further analysis is based on statistical theories whose over-riding importance is that their basis in mathematics makes their exploitation with digital computers very promising and holds out the possibility that part of the diagnostic process can eventually be taken over by the computer.

INFORMATION THEORY states that 'information' can be gained in two distinct ways: by communication in some agreed language or sign system in which a message can be assembled; and by observation - when, for example, we look at a picture or down a microscope. The exchange of information (the signal) takes place across a 'channel' connecting the patient or sender and the clinician or receiver. The amount of information transmitted by a channel/

channel can be measured and the upper limit of the amount that any channel can transmit is known as the 'channel capacity'. The channel often transmits information other than the signal (noise).

The exchange of information between doctor and patient can be viewed in this way. The data about the patient's condition can be auditory (history, heart sounds), visual (facial expression, electrocardiogram, blood slide) or tactile (pulse). They are encoded, transmitted and then detected and decoded in the clinician's brain. The occurrence of noise in this channel can be measured as 'observer variation'. This variation takes two forms: 'intra-observer error', where one clinician differs in his assessment of the same data (such as history, physical examination, blood slides) on different occasions; and 'inter-observer error', where two or more clinicians differ in their assessment of the same set of data. Information theory allows us to measure both the information provided by a piece of data and the effect of noise. It can be used to study mathematically this important aspect of the diagnostic process and to compare different methods of acquiring and assessing clinical data (whether this is provided by computer or by clinician).

Items of a patient's history, physical examination and laboratory tests all produce single pieces of information, each of which must be processed in a similar way. Thus, the question "Is your neck painful?" or the observation "Is the thyroid gland hard, firm or soft in consistency?", can both be called tests.

When the clinician has in hand the outcomes of one or more of such tests, another stage in the diagnostic process begins. He must interpret these pieces of information and assess them in the light of his factual knowledge and personal experience of the group of diseases from which he thinks his patient may be suffering. He revises his opinion about diagnosis and treatment progressively as new information becomes available.

Once again there is a well established theory of statistical inference typified by conditional probability theory and by Bayes theorem in particular which can describe this process. Bayes theorem is widely used in studies of computer-assisted diagnosis. It allows us, mathematically, to revise our opinion of a hypothesis (for instance, deciding if a patient is suffering from thyroid cancer) in the light of new information, such as the outcome of a test. The calculations are based on previous/

previous knowledge of the diseases being studied. This knowledge is summarised in the form of 'likelihoods' which predict from past surveys the relative frequencies of test outcomes in each disease.

In this way likelihoods are used to summarise quantitatively previous experience of the diseases. Presumably the clinician stores the same information derived from his own past experience of a large number of diseases in his brain in some coded form. He uses this data to revise his opinion in the light of new information by psychological mechanisms as yet unknown.

Apart from these likelihoods, the only other retrospective data used in computer-assisted diagnosis are the prior probabilities of each disease. These are the incidence rates of the diseases in the population being studied. The probabilities vary in different populations. Thus, tropical diseases are common around the equator and rare in Britain.

So by combining prior probabilities, likelihoods and test outcomes (on the patient being assessed), by means of conditional probability theory we can represent mathematically the process of inference involved in the diagnostic process. Many of the studies of computer-assisted/

assisted diagnosis are based on conditional probability theory and have been found to give an accuracy in diagnosis comparable to that of clinicians.

The third theoretical component of this analysis of the diagnostic process is statistical decision theory. This describes mathematically the decision-making process which underlies the selection and assessment of clinical information and the choice of investigations and treatments. At each step of the decision process a number of factors has to be weighed up. These include the advantages of making an immediate correct diagnosis and the consequences of making an overhasty misdiagnosis. Furthermore, the financial cost, discomfort, inconvenience and delay due to further investigations before treatment is begun must be considered.

THE PRINCIPLE of 'rationality' - one of the bases of statistical decision theory - when applied to the clinical situation asserts that any rational diagnostician acts as if he is able to measure the advantages and disadvantages of a decision in common units and make decisions (such as to continue testing or to make a diagnosis) so that the expected net advantage is as large as possible. In other words, he attempts to achieve the greatest benefit for/

for his patients at a minimum 'cost'. This view of the practice of medicine in terms of 'costs' represents the implicit value system with which clinicians manage their patients.

Since clinicians differ in their judgements of these costs, it can be said that each state of health or of illness has its 'value' for each clinician. Thus, one clinician in a chronic renal dialysis unit, where potential patients greatly outnumber the limited number of kidney machines available, may feel that only young adults should be accepted for chronic dialysis. By contrast, another clinician may feel only patients with a criminal record should be excluded and that all others should be put on a waiting list. The importance of such value systems is that they strongly influence the number of investigations and type of treatment chosen. Statistical decision theory provides a mathematically based framework for studying and measuring such values.

C H A P T E R 2 : THEORETICAL BACKGROUND

2.1 Conditional Probability Theory.

2.2 Information Theory and Sequential Models.

2.3 Non-linear Discriminant Function Models.

2.4 Decision Theory.

Theoretical Background to the Studies

The studies in this thesis can be conveniently divided into those using the disease system of non-toxic goitre, which are based on derivatives of conditional probability theory and of information theory, and those using a non-linear discriminant function method in the development of a dynamic model of thyrotoxicosis.

2.1 Conditional probability theory

Probability can be defined as a means of quantifying uncertainty and it may be argued that all probabilities are conditional in that one's degree of certainty about an event, such as a symptom or a diagnosis, is conditioned by the knowledge one already possesses about it. Conditional probabilities provide a probabilistic expression of our revision of an opinion in the light of new information.

The best known example of conditional probability is Bayes' Theorem which is used to combine new information with that already known to reach for example a differential diagnosis.

2.1.1 The use of Bayesian conditional probabilities has been summarised briefly as follows:

(1) Probabilities are orderly opinions

(2)/

(2) Statistics is concerned with the revision of opinion in the light of new information.

(3) Bayes' Theorem of probability theory is a mathematically optimal rule for such revisions of opinion.

(4) Such probabilities are usually expressed as fractions of 1, such that the total of probabilities in the group considered (for example diseases) add up to 1.

Before describing Bayes' Theorem, it is worthwhile restating two of the definitions already used in the introduction.

Since an item of history and of physical examination produces information in the same way as laboratory or other tests, it will be the convention in the studies described in this thesis to refer to all such items as tests. The response to a question, the result of an examination or investigation are thus referred to as test outcomes.

2.1.2 Bayes' Theorem can be expressed in many forms, but the commonest version is:

$$P(D/T) = \frac{P(T/D)P(D)}{P(T)} \qquad \text{(Unless } P(D) = 0 \text{ or } P(T) = 0\text{)}$$

Where D = disease, and T = test outcome, the standard notation/

notation of probability theory means that the expressions P(D/T) reads as "the probability of disease D given test outcome T"; for example, the probability of thyrotoxicosis (D) given that exophthalmos (test) is present (test outcome). Similarly P(T/D) reads "the probability of test outcome (exophthalmos is present) given the disease (thyrotoxicosis).

P(D) reads "the probability of disease D"

P(T) reads "the probability of test outcome T"

The expression P(D) is normally called the 'initial' or 'Prior' probability before the test T has been carried out, while P(D/T) is the 'final' or 'posterior' probability i.e. the revised opinion about disease D in the light of the new information provided by the test outcome T. P(T/D) is referred to as the likelihood.

2.1.3 Prior probabilities

The prior probability (P/D) represents the incidence rates of the diseases under consideration in the population being studied. Thus, in the non-toxic goitre study (Chapter 4) the frequency with which the three diseases are seen in the thyroid clinic at the Royal Infirmary, Glasgow, are Hashimoto's disease 10 per cent, simple goitre 89 per cent and carcinoma of thyroid 1 per cent; this gives the prior probabilities used in the study of (.10, .89, .01).

In the studies described here the population from which the sample is drawn is well defined, namely, a particular thyroid clinic in a particular hospital, and so the choice or prior probabilities is simple. By contrast, where a group of diseases is studied without any clear reference to a specific population, the choice of priors may be difficult. Boyle et al (1966) avoided this by choosing the "relative likelihood" form of Bayes' theorem in which the prior probabilities are effectively omitted. The effect of different priors has been discussed with reference to Cushing's syndrome by Nugent et al (1964) when comparing their own study in Salt Lake City, Utah, U.S.A, to that of colleagues in a similar study in Nashville, Tenessee. In the former group's data the prior probability of Cushing's syndrome was 0.25, while in the latter population it was 0.19. Thus, P(D) can be seen as the 'geographical' component of Bayes' theorem. Later in the study of non-toxic goitre the effect of different prior probabilities on the effectiveness of a diagnostic model are explored and discussed (4.6.3).

2.1.4 Likelihoods

Likelihoods might be said to represent quantitatively the previous experience of the diseases in question. Fig. 2.1 gives an example of likelihoods in non-toxic goitre which/

LIKELIHOODS					
No.	Test	Response	HASHIMOTO'S	SIMPLE	CANCER
25	Consistency	Firm	.9057	.5800	.4600
		Hard	.0566	.0400	.5300
		Soft	.0377	.3800	.0100 +
27	C.F. test	++	.8372	.0100	.0200
		+	.0698	.0513	.1081
		-	.0930	.9387	.8719
8	Pyramidal Lobe	Absent	.8491	.9608	.9783
		Present	.1509	.0392	.0217

FIG. 2.1. EXAMPLE OF LIKELIHOODS

which is taken from the matrix of likelihoods published by Boyle et al (1966). It shows that a 'firm' gland is much more likely to be Hashimoto's disease (.9057) than simple goitre (.5800) or thyroid carcinoma (.4600); while a 'hard' gland is more likely to be thyroid cancer (.5300) than Hashimoto's disease (.0566) or simple goitre (.0400). By contrast, the presence or absence of a pyramidal lobe is very similar in all three diseases and so this test is of little discriminating value in these three diseases.

A sequential technique to select tests on the basis of their expected usefulness in discriminating within a group of diseases is the basis of the first part of this thesis. The theoretical basis of this technique is described later (2.2).

2.1.5 Posterior probabilities

The posterior probabilities represent the revised statistical opinion in the light of the available data, i.e. the differential diagnosis. The final diagnosis is allocated to the disease with the highest posterior probability. In the study of non-toxic goitre (Chapter 4) a posterior probability of .99 which is unchanged for three consecutive diagnostic orders was chosen as the "stopping" rule for an acceptable final diagnosis. In cases/

cases where the final posterior probability is less than .99, the diagnosis may be regarded as less reliable.

The number of such uncertain diagnoses in a sample of test cases must clearly be taken into account. In the assessment of the value of a model for computer-assisted diagnosis Crooks et al (1959), Nugent et al (1964) and Anderson and Boyle (1968) have each introduced a category of uncertain or 'queried' diagnosis in which judgement is suspended until further information is available. Crooks et al (1959) using a linear discriminant model left an area in the centre of the linear scale of their diagnostic index such that a score of from -16 to +11 would indicate a diagnosis of "definite non-toxic", a score of more than +19 was "definite thyrotoxic" and the intermediate score of +11 to +18 was denoting the "equivocal range". Nugent et al (1964) in their Bayesian model for Cushing's syndrome, used a final posterior probability of .90 or more for a 'likely diagnosis' with judgement being reserved in other cases (less than .90). Anderson and Boyle (1968) used as a stopping rule that the largest posterior probability should be at least ten times that of the next largest. Other aspects of the problem of diagnostic certainty and misdiagnosis are discussed in detail later (4.6).

2.1.6 Independence of tests

The assumption of statistical independence between test is made in all studies to date of computer-assisted diagnosis by conditional probability models. This is, strictly speaking, unrealistic in practice since the various clinical, biochemical and other manifestations of a disease or a pathological process are likely to have some relationship to one another. The important point is whether these inter-relationships are weak or strong and whether there is any theoretical or practical benefit in attempting to measure them and to take such interactions into account in the model.

In many studies in the literature the assumption of independence is acknowledged, briefly discussed but disregarded in the calculations (Overall and Williams, 1961; Winkler et al, 1957). Some workers have attempted to reduce the effect of interaction on the assumption, not so far conclusively proven, that failure to take interaction into account would reduce the success of the model. Warner et al (1964) grouped together four different types of diagnoses which were known to be interdependent and allowed one positive response for one or more of them being present. Lodwick et al (1965) adopted a similar approach/

approach of grouping related tests in their model for the diagnosis of primary bone tumours, so that only one test outcome was recorded for the whole group. Nugent et al (1964) prepared contingency tables for the actual and expected coincidence of 154 possible paired combinations of 13 tests in the patients studied. The extent of interdependence was calculated on the basis of the Chi square test. Since two tests ((a) the presence or absence of ecchymoses and (b) a serum potassium of 3.6 m.eq./litre or less) were involved in the five significant associations found, these were eliminated and the remaining 11 tests used in the model.

In the sequential model of non-toxic goitre used in this study the data used is that of Boyle et al (1966). In the original preparation of the matrix of likelihoods these authors eliminated three serum tests known to be based on serum globulins and in two other serum tests (thymol turbidity and zinc sulphate turbidity) the result of only one or other of the tests was used, never both. Independence among the remaining tests was otherwise assumed.

The non-linear discriminant function model used in the second part of the thesis uses continuous data and some/

some account is taken in the calculations of interdependence among the tests used.

In practice the assumption of independence is dictated by the lack of sufficiently large samples to take account of interdependence and by the great increase in computational time needed. Only in situations where models based on the independence assumption do not function efficiently (compared with clinicians) will the more sophisticated model be necessary. In more theoretical analyses of the mathematical 'structure' of diseases, as in the thyrotoxicosis study described later, the more sophisticated models are preferable to prevent mis-interpretation of data.

2.2.1 Sequential probabilistic models and information theory

In the introductory chapter the sequential nature of clinical decision-making was described and in the preceding section of this chapter the use of Bayes' theorem in the 'static' models of computer-assisted decision-making has been discussed in detail.

In the literature to date only two forms of sequential decision models have been developed.

The first group fall into the category of logical decision trees where the outcome of tests are recorded as Yes/No/

Yes/No or equivalent classifications. Examples of such logical decision trees are those of Edwards (1970) (for the anatomical and pathological diagnosis of dysphagia), Kleinmuntz (1965) and Wortman (1966) in the neurological field, and Slack et al (1966) who used such a logical decision tree as the basis of a computer-based history taking system. Several systems of this type are reviewed in Taylor (1970c).

Such logical decision trees are rigid and it is possible to take the wrong route in the decision tree should a test outcome be uncertain or incorrect without any possibility of correcting the route in the light of later information. They are likely to make a limited contribution to the general problem of the clinical diagnosis.

The other group of sequential decision models are those based on statistical decision theory. Ginsberg and Offensend (1968) dealt with a simple decision problem about the type of biopsy selected to investigate a collapsed spinal vertebra in a 5 year old boy, while Gorry and Barnett (1968) used a limited form of statistical decision theory on the data of Warner and his associates (1965) for congenital heart disease. The role of decision theory is discussed later in this chapter (2.4) and a detailed analysis of the above studies has already been presented in the literature review (3.5).

2.2.2 The sequential probabilistic model

In clinical practice clinicians collect data on their patients in a sequence, being guided at each stage of selecting the next test by a mental estimate of the probabilities of the diseases under consideration. He chooses the test (or group of tests) which he expects to yield most information at the stage he has reached in the case being considered.

It is possible to select tests in a similar way by using a calculation which predicts the most informative test. The technique used in this study was first described by Lindley (1956) when he suggested using the information or 'entropy' measure of Shannon (1948), one of the key concepts of information theory, for statistical problems. This measure was originally used to measure mathematically the amount of information transmitted by a communication channel. Lindley suggested that this could be applied as a measure of the information provided by an experiment. This could then to used to select in a sequence those experiments which are expected to yield the greatest gain in information and continue experimentation until a pre-assigned amount of information has been obtained.

Good/

Good (1968) suggested that Lindley's use of the entropy calculation could be used to select tests in a sequence to reach a pre-assigned amount of information, namely a posterior probability of an acceptable value. Each test is then seen as an experiment.

The technique suggested by Lindley uses the posterior probabilities as calculated by Bayes' theorem for each test separately. The calculations to select the 'best test' from the available 30 tests in the non-toxic goitre study (Chapter 4) is as follows:

(1) For each test outcome the appropriate likelihood (from the table of likelihoods) is used to calculate the posterior probability by means of Bayes' theorem. Hence with, for example, 30 tests and two possible outcomes for each, we will have 60 posterior probabilities.

(2) The posterior probabilities for each outcome of each test are combined by Shannon's information or 'entropy' measure to arrive at a value for the expected gain of information for each test.

(3) The test with the lowest entropy value (or conversely the highest information value) is selected as that which is likeliest to yield the greatest gain of information whatever the test outcome in the patient in question proves to be.

SUMMARY OF SEQUENTIAL PROBABILISTIC METHOD

1. From likelihood and prior probabilities use Bayes' theorem to calculate posterior probabilities.

2. Posterior probabilities are combined by Shannon's 'entropy' measure to arrive at value for 'expected gain of information' for each test.

3. Select test with lowest 'entropy' value as the one which is likeliest to yield greatest gain of information.

4. Test outcome from selected test entered at terminal.

5. From all the posterior probabilities calculated in '1', the posterior probability corresponding to the test outcome is printed out at the terminal. They also become the prior probabilities for the next cycle.

FIG. 2.2. SUMMARY OF THE SEQUENTIAL TECHNIQUE

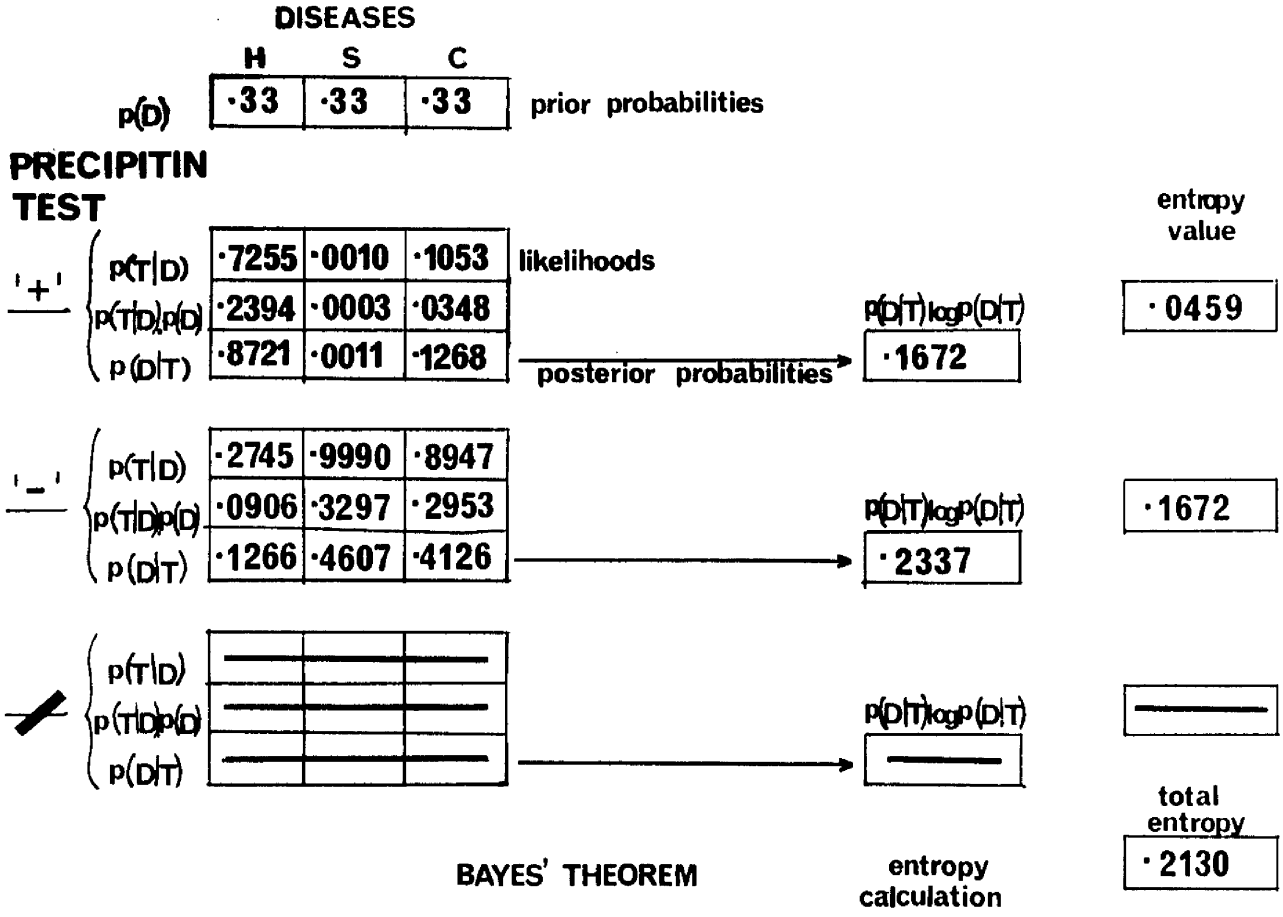


FIG. 2.3. EXAMPLE OF THE CALCULATION OF AN ENTROPY VALUE

The calculations just described are summarised in Fig. 2.2, while in Fig. 2.3 an 'entropy' value is calculated for each test outcome and the total entropy per test is shown.

(4) In the program used in the study of non-toxic goitre the 'best test' is selected by the above means. The test outcome is entered at the terminal and the posterior probability (already calculated for each test outcome) appropriate to the outcome is used as the new prior probability for the next cycle. This posterior probability is printed out at the terminal and is used to plot the graphs shown later (Fig. 4.8 - 4.11).

The value of such a probabilistic sequential decision tree is that it more closely resembles the actual practice of clinicians than the 'static' models so far used. The logical decision trees are much more rigid and do not allow the effect of erroneous test outcomes to be overcome later as in the diagnostic sequential probabilistic model.

2.3 Non-linear discriminant function model

The normal discriminant function technique attempts to separate two or more groups of data by a straight line (in a two dimensional model) or its equivalent in higher dimensions. In some cases no such line can be drawn which will/

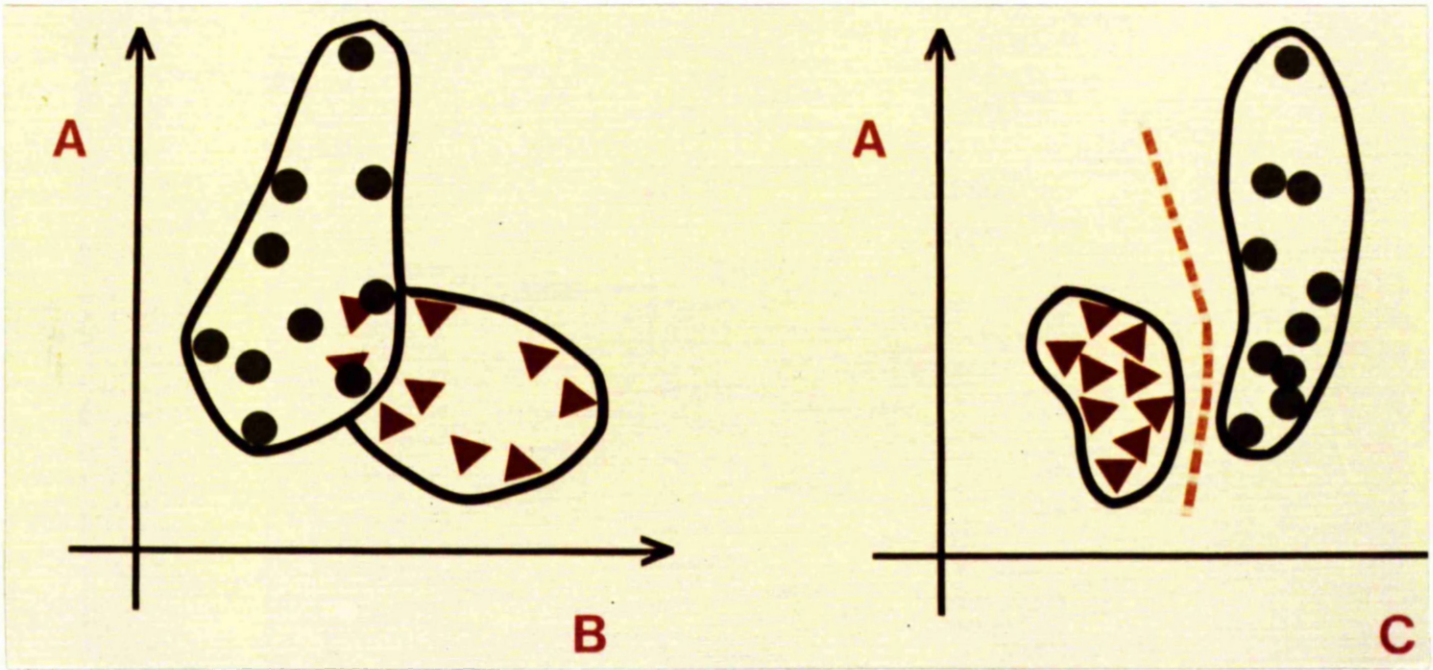


FIG. 2.4. (a) SEPARATION IS NOT POSSIBLE IN TWO DIMENSION BUT IS POSSIBLE IN THREE.

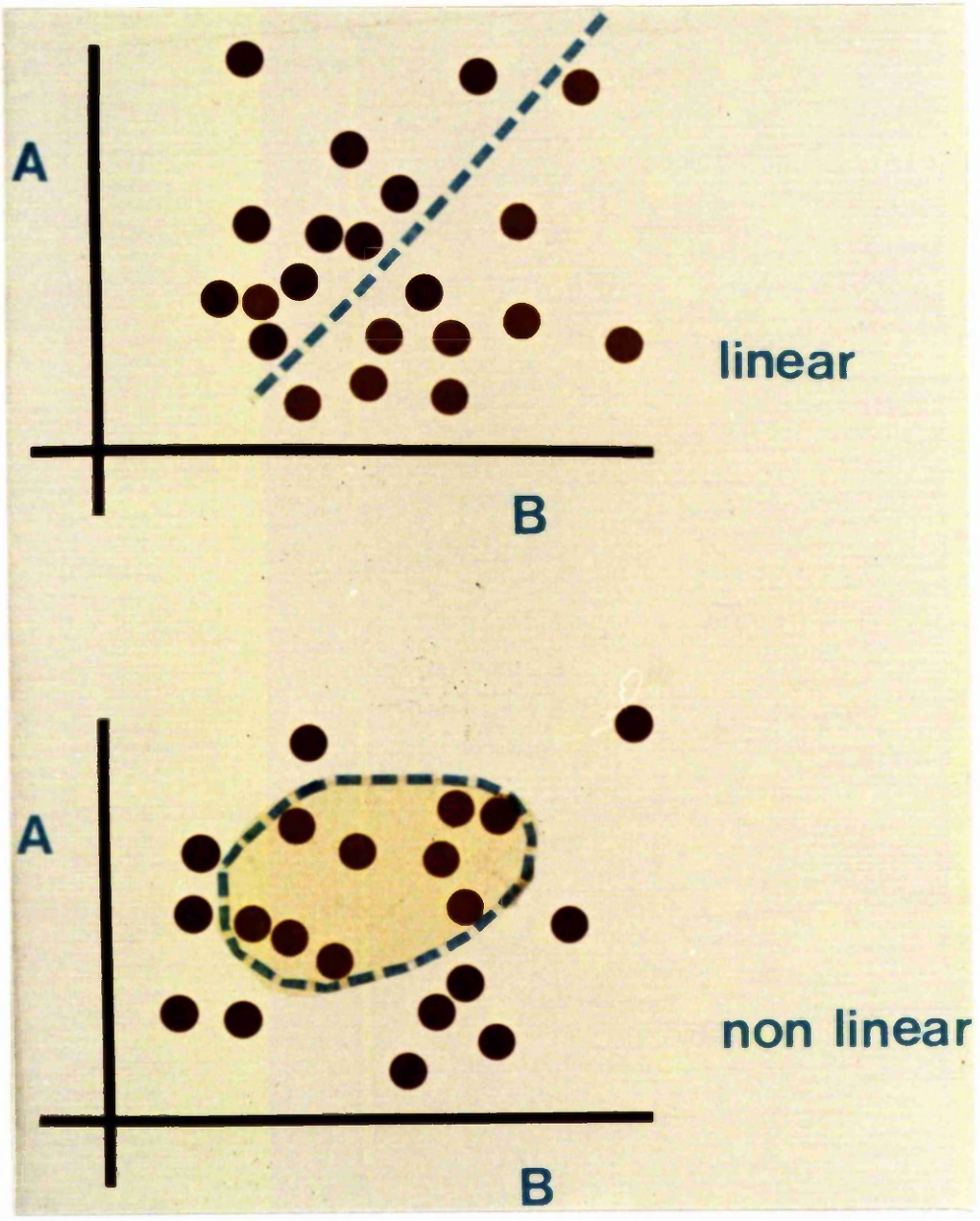
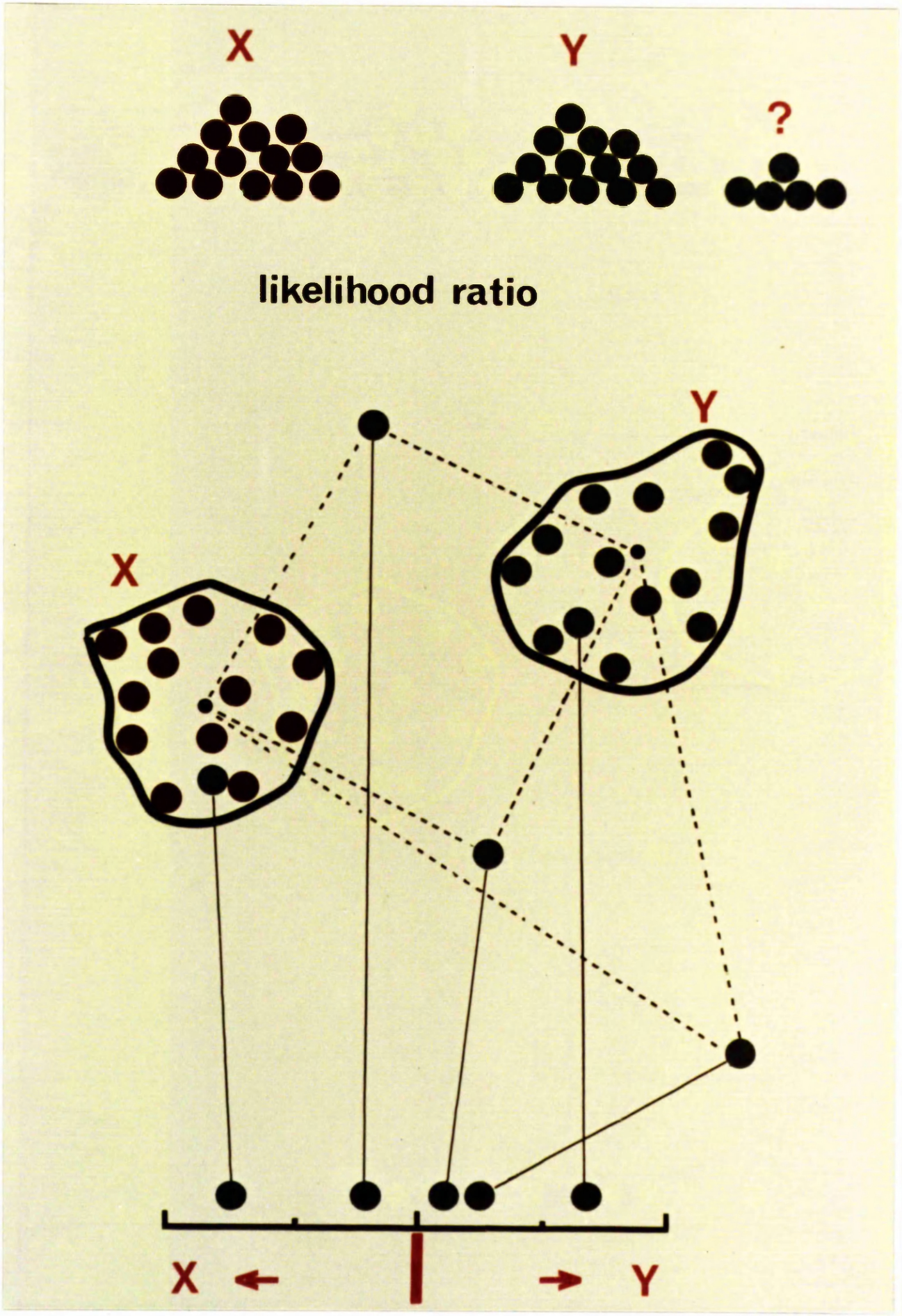


FIG. 2.4. (b) SEPARATION IS NOT POSSIBLE LINEARLY BUT IS POSSIBLE IN A NON-LINEAR MODEL.

FIG. 2.4. (c) NON-LINEAR DISCRIMINANT FUNCTION MODEL: THE KNOWN GROUPS X AND Y ARE CIRCUMSCRIBED AND THE RELATIONSHIP OF THE UNKNOWN CASES TO EACH CLUSTER IS CALCULATED ON THE BASIS OF THE RELATIVE DISTANCE OF EACH UNKNOWN CASE TO THE CENTRE OF EACH CLUSTER. THE LIKELIHOOD RATIO IS REPRESENTED ON THE SCALE IN THE LOWER PART OF THE DIAGRAM.


X
Y
?
likelihood ratio
Y
X
X
Y

will separate the groups. A non-linear model (Fig.2.4) uses an elipsoid to circumscribe each group in the following manner (Aitchison, 1971, Personal communication). Suppose that a number of patients have been diagnosed into two classes and that for each patient the results of t diagnostic tests are available.

The technique defines the two groups of patients as two clusters in multidimensional space. A measure known as the quadratic form for each case is computed which represents the relationship of that particular case to the cluster to which it belongs, and a similar quadratic form is calculated for the relationship (in multidimensional space) of the same case to the other cluster(s). The value that the quadratic form takes in any case is determined by the test outcomes in the case in question.

The basis of the technique can be stated in a more precise mathematical form in the following way:

Each patient is represented by the outcomes in his case of the set of tests chosen to investigate the diseases in question. The cases are already in a diagnostic category or class. The set of test outcomes in any one case is denoted as the vector of outcomes in t diagnostic tests.

Let/

Let C_1 and C_2 denote the two classes of vectors of test outcomes (for example in thyrotoxicosis the classes are "cured" and "relapsed").

Let $\underline{\bar{x}}_i$ and $\underline{S}_i$ be the vector of t means and the covariance matrix of the ith group of patients (i = 1,2). For any t-dimensional vector $\underline{x}$ (test outcomes in a single patient), define the two quadratic forms:

$$d_i(x) = (x-\bar{x}_i)' S_i^{-1} (x-\bar{x}_i),$$

and let

$$ki = \max_{x \in C_i} d_i(x)$$

(that is, let ki = the maximum value that the quadratic form $d_i(\underline{x})$ takes in class i)

then $Q_i = (x : d_i(x) \leq k_i)$,

define the circumscribing quadrics of the two classes.

The circumscribing quadrics are fitted by the computer program to each group in the form of "ellipsoids" taking the 'shapes' of these clusters and of such a size that they just contain all the points of their cluster.

The degree of separation of the two classes can then be investigated in two stages -

(1) by geometrical considerations

(2) by likelihood ratio.

(1)/

(1) Any x in C_1 such that $d_2(x) \geq k_2$ and any x in C_2 such that $d_1(x) > k_1$ is said to be separated by geometrical considerations alone. Any x in C_1 or C_2 not so separated lies in $Q_1 \cap Q_2$, the intersection of the two quadrics.

(2) Any x in C_1 or C_2, not separated by (1), may be investigated by consideration of the likelihood ratio

$$\Lambda(x) = \left(\frac{\det S_2}{\det S_1}\right)^{\frac{1}{2}} \exp\left\{-\tfrac{1}{2}\left[d_1(x) - d_2(x)\right]\right\},$$

where det (.) denotes determinant. Our rule is to regard x in C_1 as separated if $\Lambda(x) > 1$ and, similarly x in C_2 as separated if $\Lambda(x) < 1$.

For the diagnosis of a new patient with vector x of test results we compute $d_1(x)$ and $d_2(x)$. If $d_1(x) \leq k_1$ and $d_2(x) > k_2$ then x lies in Q_1 but outside Q_2 and we diagnose as in first class ('cure'). If $d_1(x) > k_1$ and $d_2(x) \leq k_2$ then x lies in Q_2 but outside Q_1 and we diagnose as in second class ('relapse'). If $d_1(x) \leq k_1$ and $d_2(x) \leq k_2$ the patient is in both Q_1 and Q_2; we then compute $\Lambda(x)$ and diagnose as in first or second group according as $\Lambda(x) > 1$ or < 1. If $d_1(x) > k_2$ we say that the new patient is 'outside previous experience'; in this case we again compute $\Lambda(x)$ and diagnose as above.

The following points should be noted in addition:

1. The number of patients in each class must be at least/

least t+1.

2. Separate covariance matrices are computed for the two classes and previous experience with the technique suggest that the covariance matrices are sufficiently different to make linear discriminant analysis of inferior quality.
3. Similar experience suggests that the likelihood ratio will be more useful in practice than the geometric method.

2.4 Statistical decision theory

The concepts of decision theory are not immediately relevant to the actual experiments in this thesis but many of the problems of diagnosis and computer-assisted diagnosis can be illuminated if viewed in terms of decision-making.

In the introductory chapter (1.3) the importance of the value system which guides the clinician in his choice of investigations and treatment was emphasised. The concept of financial and non-financial costs was also put forward. The 'principle of rationality' was shown to assert that any rational diagnostician acts as if he is able to weigh up the advantages and disadvantages of a decision in common units and make decisions so that the/

the expected net advantage is as large as possible. In this way he attempts to achieve the greatest benefit for his patients at minimum cost.

The concepts of decision theory were put forward in the earliest writings on computer-assisted diagnosis (Ledley and Lusted, 1959) but so far have had little practical relevance in studies to date. Lusted (1968) discusses its value in calculating the choice of optimal treatment and Aitchison (1970) proposes a technique for the calculation of costs which clinicians attach to the treatments which they choose.

Since no study involving statistical decision theory is included in the thesis no further elaboration of the theory is needed, apart from the general principles covered in this section and in the literature review.

C H A P T E R 3 : HISTORICAL REVIEW OF COMPUTER-ASSISTED DIAGNOSIS

3.1 Introduction: early studies

Computer-assisted diagnosis has developed over the last 15 years from studies based on symbolic logic or discriminant function analysis (often deducing or calculating differential diagnoses without actually using a digital computer) to the present day emphasis on conditional probability theory.

Several of the earliest studies used discriminant function analysis to calculate the diagnosis in liver disease (Zieve and Hill, 1955), in thyroid disease (Crooks et al, 1959), in lung cancer (Hollingsworth, 1959) and to estimate prognosis in pulmonary tuberculosis (Oyama and Tatsuoka, 1956). Concepts of symbolic logic were used in several early studies to deduce logically a differentiation diagnosis. Nash (1954) developed a 'logical slide rule' for differential diagnosis among 336 common diseases, while Ledley (1956) and Lipkin and Hardy (1958) used a similar logical approach in haematology by means of marginal punched cards.

However, the most important early paper in the whole field of computer-assisted diagnosis was undoubtedly that of Ledley and Lusted (1959) which laid the foundation in symbolic logic, probability theory and statistical decision theory for all the significant research since then.

The general picture that has emerged since 1959 is that of active research in many specialist areas, in particular thyroid disease, cardiology and haematology, with a strong emphasis on Bayesian conditional probability models. However, in the last few years such studies have decreased and no serious attempt has been made to apply the techniques already available to 'real-life' clinical situations. This is probably due in part to "lack of a sufficient number of people endowed with the necessary cross disciplinary background in both computer science and medicine" (Ledley, 1969).

The main reason is almost certainly that the diagnostic problems so far chosen for investigation do not quite fit the day-to-day clinical material seen by clinicians. All studies to date have assumed that only one disease is present (with the possible exception of Warner et al (1961) in their study of congenital heart disease) and have largely ignored the importance of psychological disorders which very commonly co-exist with or simulate organic disease (Taylor, 1969). The author has proposed a two-tier out-patient system in which the first screening tier is based on a system of computer-assisted diagnosis (Taylor, 1970) using the sequential decision model described in the previous/

previous chapter (2.1). An important feature of this system will be the concentration on taking account of dual pathology (for example, simple goitre with an anxiety state) and of psychological diseases.

In the following review of progress between the early studies just described and the present the headings indicate the technique used. The most important techniques are those of symbolic logic, logical decision trees, discriminant function analysis, conditional probability theory and statistical decision theory. A brief discussion of the various techniques of numerical taxonomy used in the closely related field of disease classification is included.

3.2 Symbolic logic and logical decision trees

The paper of Nash (1954) proposing the use of a 'logical slide rule' was probably the earliest paper in this field. The slide rule had a reference set of 336 common diseases listed along one edge. Strips corresponding to the symptoms, signs and test results in an unknown case were inserted into the 'Logoscope' and the differential diagnosis is 'read-off' where a complete line is seen across the whole of the slide rule. More recently Nash (1960)

1960,1963) has applied this technique to a wide variety of clinical problems and advocated the use of symbolic logic in clarifying our ideas about the process of diagnosis.

Ledley (1956) and Lipkin and Hardy (1958) both advocated the use of marginal punch cards in arriving at a logical differential diagnosis. Each disease is represented by a card with all the disease features recorded as marginal punches. By mechanically extracting the disease cards with all the features of the test patient's case, a differential diagnosis is automatically reached. Since there is one disease per card (i.e. multiple pathology is not considered) and sorting is done only on the positive presence of a feature (i.e. no significance is attached to the absence of a feature) the system is of limited value. Lipkin and Hardy (1958) applied this method with limited success to 26 different blood diseases. A similar system for use in ophthalmology has been developed by Paycha (1955, 1960). More recently Barker and Bishop (1970) have used a computer program which scans patient's records for combinations of symptoms and signs which are associated with a high frequency of hypothyroidism.

Another/

Another use of symbolic logic has been its application to problems of disease classification and definition. Feinstein (1963, 1964, 1966) has written extensively in this area with particular reference to the definition and prognosis of acute rheumatic fever.

More recently the use of logical decision trees has combined the concepts of symbolic logic with the use of on-line inter-active computer terminals. By a series of questions and answers at a terminal the patient or clinician is led logically to a diagnosis.

The most important application of such logical decision-making is with history taking terminals connected to a computer. In Great Britain, Edwards (1970) has developed a logical decision tree for the diagnosis of dysphagia. This is based on his refinement over many years of an elegant anatomical, functional and pathological analyses of the mechanisms of dysphagia. He has developed this for use with an on-line microfilm type computer terminal connected to a remote computer for automated history-taking. Fries et al (1970) has used a similar logical model in the diagnosis of several types of arthritis.

Warner Slack and his group in the University of Wisconsin/

Wisconsin (Slack et al, 1966, 1968) have used a L.I.N.C.8 computer with its small visual display for general history taking. More recently Mayne et al (1968) at the Mayo Clinic have experimented with a rather complex microfilm terminal (with a light pen) for general history taking. Grossman et al (1968) at the Massachussets General Hospital have used a simple teletype for history taking in which the patient is led through over 200 general history questions by a simple logical decision tree. The Mayo Clinic system involves about 300 possible questions and the Slack system has about 400 questions.

All of these systems are reviewed in Taylor (1970) including a commercial history taking system which is already being marketed by a subsidiary of an American drug firm (Medidata Inc., Waltham, Massachussets, U.S.A.). This system is based on a PDP 8I computer and uses as its data base a well established medical questionnaire the Cornell Medical Index (Bordman et al, 1949).

The disadvantages of such logical decision trees are their rigidity and the fact that erroneous data may lead irreversibly down a quite misleading path in the decision tree. The sequential probabilistic model described in the first part of this thesis has none of these disadvantages.

3.3 Discriminant function analysis models

Studies using a discriminant function model were, as we have seen above, among the earliest in the field of calculated or computer-assisted diagnosis.

Zieve and Hill (1955) used this technique to assess 11 liver function tests in normal and cirrhotic patients and showed that 4 of the tests could differentiate between normal and abnormal liver function as effectively as the whole group. Oyama and Tatsuoka (1956) used 13 items of information to produce a prognostic score for pulmonary tuberculosis which was 75 per cent correct in its prediction.

Among the best known studies with a discriminant function model is that of Crooks et al (1959) in the development of their clinical diagnostic index for thyrotoxicosis. A complimentary index for the diagnosis of hypothyroidism was later produced by the same group (Billewicz et al, 1969). Gurney et al (1970) have modified the original thyrotoxicosis index to take account of psychological illnesses. All of these thyroid studies based on linear discriminant functions will be covered in more detail in the section of this review dealing with thyroid disease.

Hollingsworth/

Hollingsworth (1959) investigated the early diagnosis of bronchial carcinoma by linear discriminant function and produced a linear model which was 85 per cent efficient with the 200 cases used. More recently Hughes et al (1963) and Norris et al (1969) have used a similar approach to survival and morbidity after myocardial infarction. Norris et al (1969, 1970) produced their prognostic index to help in the selection and management of patients in a coronary care unit. Neurath et al (1969) have used a similar approach to pre-operative diagnosis before pelvic surgery. Ferris et al (1970) recently used a non-linear discriminant function model (used in Chapter 7 of this thesis) in the differential diagnosis of Conn's syndrome.

3.4 Conditional probability models

By far the most frequently used models in studies of computer assisted diagnosis are those based on Bayes' theorem (2.1) or its derivatives.

The earliest paper on the use of conditional probability models in computer-assisted diagnosis is that of Ledley and Lusted (1959). In this paper the authors laid the broad foundation in symbolic logic, conditional probability theory and statistical decision theory on which most of the subsequent work in this area is based.

In a series of papers since then (Ledley and Lusted, 1960; Ledley and Lusted, 1962; Ledley, 1966) both authors have developed further the general theoretical framework in the field and have, in particular, extended the theoretical analysis of the use of decision theory in treatment planning. Ledley (1959) in particular has developed a theoretical basis for automatic pattern recognition and produced the first system for automatic pattern recognition of chromosomes (Ledley, 1966, 1969). More recently Lusted (1968) has published a monograph on medical decision-making which assembles and analyses critically most of the work in the field of computer-assisted diagnosis. The monograph also includes a detailed analysis of the important topic of observer error.

Most other studies in the field of conditional probability models have applied Bayes' theorem to differing populations of patients. Some have confined themselves to applying the technique accompanied by some comments on its value in research and in clinical practice in their own speciality. Others have attempted to resolve some of the problems involved in the use of Bayesian models such as the independence assumption, the effect of prior probabilities and the importance of non-financial costs. All/

All of these topics are discussed in detail in the previous chapter and in the body of this thesis.

The review of conditional probability studies is perhaps best made under the headings of individual specialities.

In the field of Haematology Lipkin (1964) has used the likelihood ratio (i.e. the relative odds on each diagnosis) in the differential diagnosis of 26 blood diseases. In this study the computer "matched" the test case to the most likely disease to produce a list of differential diagnoses and to suggest further tests.

In the field of Gastro-enterology Rinaldo et al (1963) and Scheinok and Rinaldo (1968) have used a conditional probability model to analyse 11 items characterising upper abdominal pain and have compared the Bayesian model with a discriminant function model. The data were carefully defined in both papers and even with 8 items from the history alone the computer program was over 69 per cent accurate in hiatus hernia, gallstones and duodenal ulcer patients. An important feature of these studies was that the aim was to predict the radiographic findings and not an operative or histological diagnosis. In the comparison of the Bayesian and discriminant function models the/

the former was more accurate for duodenal ulcer, while the latter was best for gastric ulcers. Wilson et al (1965) used a conditional probability model in developing criteria for the radiological differentiation of benign and malignant gastric ulcers. From an original list of 70 items, Bayes' theorem was used to identify a final list of 31 statistically significant variables.

The most important studies in the field of Cardiology have been those of Warner et al (1961). From a large amount of clinical and laboratory data they have developed a Bayesian model for 53 items on 35 types of congenital heart disease. In a series of studies (Toronto et al, 1963; Warner et al, 1964; Warner et al, 1965) many important problems including statistical independence of data and dual pathology have been thoughtfully analysed. Templeton et al (1966) and Reale et al (1968) developed a similar model for the radiological diagnosis of congenital heart disease, as have Bruce and his co-workers (Bruce, 1963; Bruce et al, 1961, 1966) in the field of valvular and ischaemic heart disease.

In other areas of clinical medicine Lodwick et al (1965) had an accuracy rate of 80 per cent in a study of primary bone tumours, while Overall and Gorham (1963) dealt/

dealt in detail with psychiatric diagnosis. The application of a Bayesian model to Cushing's syndrome by Nugent et al (1964) tackled many important problems, including the effect of prior probabilities, independence of data and the comparison of the data used in the study with identical data collected in another centre.

3.5 Statistical decision theory

Apart from the theoretical studies of Ledley (1966, 1969), Lusted (1968) and Ledley and Lusted (1959, 1962) there are very few studies of decision theory in a medical context. Gorry and Barnet (1968) developed a theoretical sequential decision model based on artificial costs in which a treatment was chosen by a path of 'minimal cost'. Ginsberg and Offensend (1968) have used a decision theory framework to analyse the simple decision problem of choosing a method of biopsy in a child with a collapsed spinal vertebrae. This decision system has four possible diagnoses (bone infection, bone cancer, histiocytosis and rheumatoid nodule), four treatments and four possible outcomes (cure, kyphosis, paralysis or death). A detailed analysis of the problem including decision trees is included in this interesting theoretical analysis.

Two/

Two recent studies of treatment choice have leaned heavily on decision theory (Aitchison, 1970; Card and Good, 1970). The former deals in considerable detail with utility estimation in general and in medicine in particular, and proposes a technique for calculating utility functions from the decision-making behaviour of clinicians which is being used in a study of decision-making in thyrotoxicosis by the writer.

3.6 Critical analysis of computer-assisted diagnosis

Sterling and Pollack (1966) critically reviewed the approach to computer-assisted diagnosis. They identified three types of diagnostic problems; the first category includes the situation where the clinician starts with no prior knowledge of the patient and the authors suggest the field is too wide for any computer system. Experience with general history-taking systems (Slack et al, 1968; Grossman et al, 1968) suggests this is too pessimistic a view.

The second category is that of differential diagnosis such as in a specialist clinic. It is in this category that they felt computer-assisted diagnosis was most promising.

The third category is in electrocardiographic analysis, processing/

processing of radioisotope scans and radiation treatment planning and success in this area is already apparent.

The most difficult problem of all, according to these authors, is use of computers in decisions where value judgements are involved. The studies described in an earlier section (3.5) of this review are elementary attempts at this problem.

3.7 Computer-assisted disease classification

The final part of this review deals with an area which, like observer error studies, is closely related to the topic of this thesis. Because of the similarity of the techniques used in the studies of thyrotoxicosis in Chapter 7 to traditional methods of numerical taxonomy, these will be dealt with in some detail. The most important techniques used in this area are symbolic logic and numerical taxonomy.

One of the most active workers in the field of disease classification is Feinstein (1963, 1964, 1966, 1970), who makes extensive use of the concepts of symbolic logic and in particular Venn diagrams, which portray the logical relationships between groups of symptoms by means of overlapping circles. Such techniques are useful in clarifying/

clarifying ideas about diseases and Feinstein et al (1964) have made very effective use of such techniques in their studies of the variations in the manifestations and in the prognosis of acute rheumatic fever.

In the more complex problems of classification, the group of techniques known as numerical taxonomy have been used. These were originally developed for use on bacteriological and botanical classifications and during the last decade over 30 methods have been proposed to measure the inter-relationship between organisms, plants or patients (Wishart, 1969). The subject of numerical taxonomy has been extensively reviewed in an important book by Sokal and Sneath (1963).

The general objective of numerical taxonomy (or 'cluster analysis' as it is often known) is to partition a population of individuals into "meaningful" or "useful" classes. The computer programs used in these studies scan the set of data representing a large population of patients with a set of related diseases, to detect similarities between pairs of patients. This similarity is measured by means of a similarity co-efficient and the computer 'links' together in pairs into clusters those patients who are most similar. Over 30 different techniques, based on/

on different co-efficients of similarity, are now in use (Sokal and Sneath, 1963; Ward, 1963; Wishart, 1969).

Numerical taxonomy has been applied with success to the leukaemias (Hayhoe et al, 1964; Knox, 1964; Mantel, 1967), pyelonephritis (Zinnser et al, 1962), cardiology (Manning and Watson, 1966) and liver disease (Fraser and Baron, 1968). The latter authors have reviewed the application of numerical taxonomy to medicine (Baron and Fraser, 1965).

Techniques of numerical taxonomy differ radically from the non-linear model for cluster definition used in the second half of this thesis. In the non-linear model the clusters are specified at the outset of the study and a mathematical model is fitted to each cluster. Unknown cases are then diagnosed on the basis of their relative similarity to these clusters.

In numerical taxonomy, as we have seen, the clusters 'emerge' in the course of the analysis.

3.8 Studies of thyroid disease

The earliest study in thyroid disease was the attempt by Schultz and Zieve (1956) to produce a remission after a single dose of radioactive iodine in thyrotoxic patients. This/

This study used a score obtained by allocating weighted values at intervals after therapy to the clinical state, the thyroid uptake of ^{131}I, the basal metabolic rate and the level of serum cholesterol.

The most important early study was that of Crooks et al (1959) who used the symptom analysis of Wayne (1954) to identify the most discriminating items in patients suspected of thyrotoxicosis. A linear discriminant function analysis was used to produce weighting factors for the chosen symptoms and signs. A total score of $\geq +19$ was taken as 'definitely thyrotoxic', $\leq +11$ was 'definite non-toxic', while the intermediate zone of +11 to +19 was designated the 'equivocal range' where clinical judgement should be reserved. A success rate of 85 per cent in 171 test cases was obtained. A decade later the same group produced a similar index for hypothyroidism (Billewicz et al, 1969). This is now used for the follow-up of treated thyrotoxic patients (Hedley, 1970) while Gurney et al (1970) have modified the thyrotoxicosis index to take some account of psychological disorders.

The best known conditional probability model in thyroid disease is that of Overall and Williams (1961) which/

which was later developed further by Fitzgerald, Overall and Williams (1966). In this study the diagnosis was made between three types of clinical thyroid status, i.e. hypothyroid, euthyroid or hyperthyroid, and patients were allocated to these classes on the basis not of laboratory tests but of response to treatment after one year's observation. In this study the Bayesian model was 96 per cent correct.

This same model, including the original data, was later used by Winkler et al (1967). This latter group compared in detail their own data with that of Fitzgerald et al (1966) and made important observations on the effect of the definition of symptoms (in particular 'lethargy' where there was a tenfold difference in incidence between the populations). The model was 91 per cent accurate on test cases provided by Winkler et al.

More recently Grover and Gordon (1970) in Israel have used the same Fitzgerald et al (1966) program and data on 1,000 cases from Israeli centres with an accuracy of 98 per cent.

The study from which the first part of this thesis is derived is that of Boyle et al (1966) in the simple three/

three disease system of non-toxic goitre where the diagnosis rests between Hashimoto's disease, simple goitre and thyroid cancer (4.2). Two probabilistic models were studied, Bayes' theorem and the relative likelihood (where prior probabilities are assumed to be equal). The latter method was found to be superior because no cases of thyroid cancer were misdiagnosed.

C H A P T E R 4 : A SEQUENTIAL PROBABILISTIC MODEL

4.0 Summary

A sequential probabilistic model of the diagnostic process has been developed which is based on a combination of Bayes' theorem and a 'minimal entropy' calculation derived from information theory. It has been applied to the simple three disease systems of non-toxic goitre. The original model based on 155 cases has been shown to have an accuracy of between 87 per cent and 93 per cent, depending on the prior probabilities, when tested with 60 cases derived from the same clinic population.

Three experiments with different prior probabilities were conducted and the most effective model was found to be that using equiprobability, thus confirming other similar studies elsewhere and suggesting that prior probabilities might be safely ignored in future studies.

The factors including financial and non-financial costs, redundancy of tests and methods of partitioning continuous data which influence the choice of a diagnostic model are discussed in some detail, along with several illustrative examples.

The great variation in discriminating power among tests has been shown and the value of a sequential model in highlighting such variability is indicated. The superiority of some clinical tests over expensive laboratory/

laboratory tests is also clear when the sequential model is used.

Finally, the importance of the on-line diagnostic technique developed for this study is emphasised and its practicality for future larger decision systems is briefly discussed.

4.1 Introduction

Almost all examples of computer-assisted diagnosis to date are based on the 'static' model described earlier (Fig.1.1). Apart from logical decision trees (2.2), the only other sequential model is that developed by Gorry and Barnet (1968) which is based on a limited application of statistical decision theory (2.5).

In clinical practice, clinicians collect data in a sequence, being guided at each stage in the selection of the next test by a mental estimate of the probabilities of the diseases under consideration. He selects the test (or group of tests) which he expects to yield the most information at the particular stage of the diagnostic process which has been reached.

4.2 Statistical method

In this study the tests to be used are selected in a sequence by the 'minimal entropy' calculation (Lindley, 1956/

1956; Good, 1968) described in detail previously (2.2).

Statistical independence is assumed between the test outcomes in the study since the original sample was too small to measure their interdependence and to use it in the calculations.

In the three experiments in this study the prior probabilities were initially set at equiprobability (.33, .33, .33); in the second experiment the prior probabilities used were those of the clinic population from which the cases were drawn (.10, .89, .01). In the third experiment they corresponded to the distribution of diagnoses in the 60 cases used in all three experiments. These were 34 cases of Hashimoto's disease, 19 of simple goitre and 7 of thyroid carcinoma (.566, .317, .107).

4.3 Clinical data

The data used in the calculations are derived from that of Boyle et al (1966) based on a survey of non-toxic goitre in two thyroid clinics in Glasgow.

In a patient with a visibly enlarged thyroid gland who is not suffering from thyrotoxicosis (i.e. with non-toxic goitre), the differential diagnosis rests among Hashimoto's disease, simple goitre and carcinoma of thyroid.

In/

HISTORY

3 - Discomfort
9 - Pain in goitre
10 - Hoarseness
11 - Dysphagia
12 - Choking or tightness
13 - Cough or stridor
15 - Recent increase in size
23 - Duration
30 - Age

EXAMINATION

6 - Fixation to tissues
7 - Cervical lymph glands
8 - Pyramidal lobe
16 - Nodular or diffuse
24 - Estimated size of gland
25 - Consistency
26 - Clinical status

SPECIAL EXAMINATION

4 - Tracheal deviation (or compression on X-ray)
5 - Laryngeal palsy (indirect laryngoscopy)

SERUM TESTS

1 - Precipitin test
2 - Serum globulins
18 - Gammaglobulin
19 - ESR
27 - CF Test
28 - Thymol turbidity
29 - Zinc sulphate turbidity

SPECIAL TESTS

14 - $KClO_4$ discharge
17 - B.E.^{131}I
20 - 24 hour thyroidal ^{131}I
21 - $PB^{127}I$
22 - 48 hours $PB^{131}I$

FIG. 4.1. TABLE OF TESTS USED IN STUDY.

In the original study by Boyle et al (1966), thirty items of information ('tests') were used (Fig. 4.1).

The likelihoods (Fig.2.1) used in the calculations are based on 155 cases; 53 patients with Hashimoto's disease, 51 with simple goitre and 51 with thyroid carcinoma. All patients had been referred to the thyroid clinics in the Royal and Western Infirmaries, Glasgow.

The model was tested with 60 new cases derived from the same clinic population as the original data. In each patient both in the original data and in the test cases the diagnosis was either established histologically (in about 70 per cent of cases) or by agreement among physicians well experienced in thyroid diseases.

4.4. Programming technique

In studies where the 'static' view of the diagnostic process is assumed, Bayes' theorem (2.1) is applied to all the available data before calculating the final posterior probability. This was the technique used by Boyle et al (1966) and many others (3.4).

In the present study the author wrote the original computer program in Algol 60 for the I.C.L. K.D.F.9 at the Computing Service Department, Glasgow University.

The/

The original version was an 'off-line' program where the data on the test patients were stored in the body of the program as a data-file. Test outcomes on a patient being diagnosed by the program were 'read off' this file during the running of the program.

The program, using the 'minimal entropy' technique (2.2), calculates which test is expected to be the most informative at each diagnostic cycle. When this test selection procedure in the program is completed, the test outcome on the particular test selected is read off the data file on the program, the posterior probabilities appropriate to this test outcome are printed out and are used as the new prior probabilities for the next cycle. The 'off-line' version of the program written in Algol 60 is shown in Appendix A.

A typical example of the output from this version of the program is shown in Appendix A, with a simplified version of the same program written in Fortran IV for the I.C.L. 1...F.9 computer.

The main value of the sequential technique lies in its resemblence to the sequential approach to decision-making used by clinicians. In order to make it feasible for/

FIG. 4.2. TELETYPE TERMINAL USED IN THE STUDY

for a computer based diagnostic model to be used in actual practice, we must provide the computing facility in the hospital area (e.g. outpatient clinic) where decisions are made.

This type of facility is possible using an 'on-line' computer link to a computer at some distance from the clinic. The teletype terminal (Fig.4.2) is connected to the computer by a switching device, called a G.P.O. 'modem', over the public telephone circuit. Access to the computer is obtained by dialling a number using an ex-directory line.

The original 'on-line' program was written by the author in Algol 60 for the commercial 'time-sharing' system of General Electric Information Services (G.E.I.S.) and is also shown in Appendix A. A similar version of this 'on-line' program was written by the author for the on-line system (Cotan 3) provided by Glasgow University Computing Service on their I.C.L. K.D.F.9 computer.

The task of designing, writing, testing and developing the set of programs just described represents about 12-14 months' work on the part of the author.

When the terminal is in use, the program selects after a command signal from the teletype keyboard the

'best/

```
CARDS      14:27    GEIS G 01/04/70

IN SEG1
IN .FIRST
IN SEG2
IN .FIRST


THIS IS A DEMONSTRATION OF INTER-ACTIVE DIAGNOSIS.

WHEN A QUESTION IS PRINTED OUT, PLEASE TYPE THE RESULT AS GIVEN ON THE
CASE RECORD AND THE NEW PROBABILITIES WILL BE PRINTED.
IF YOU DECIDE THAT NO MORE TESTS ARE REQUIRED, TYPE '0' AS THE RESPONSE
TO THE NEXT QUESTION.
```

		HASHIMOTO'S	SIMPLE	CANCER
THE PRIOR PROBABILITIES ARE :		.10000	.89000	.01000
C.F. TEST	(27)? 1			
THE PROBABILITIES ARE NOW :		.90196	.09588	.00215
P.B. 127 I	(21)? 2			
THE PROBABILITIES ARE NOW :		.87137	.12442	.00421
PRECIPITIN TEST	(1)? 2			
THE PROBABILITIES ARE NOW :		.65129	.33845	.01026
THYMOL TURBIDITY	(28)? 2			
THE PROBABILITIES ARE NOW :		.81618	.18232	.00150
P.B. 131 I AT 48 HOURS	(22)? 3			
THE PROBABILITIES ARE NOW :		.99673	.00232	.00095
CONSISTENCY	(25)? 1			
THE PROBABILITIES ARE NOW :		.99847	.00149	.00005
GAMMAGLOBULIN	(18)? 1			
THE PROBABILITIES ARE NOW :		.99588	.00400	.00012
E.S.R.	(19)? 2			
THE PROBABILITIES ARE NOW :		.99995	.00000	.00005
FIXATION TO TISSUES	(6)? 0			

```
THIS COMPLETES THE EXAMINATION.

USED     53.00 SEC.
```

FIG. 4.3. TYPICAL 'ON-LINE' PRINT OUT FROM THE TELETYPE TERMINAL.

'best' test from the total of 30 shown in Fig.4.1. The terminal prints out the name of the tests and awaits entry by the physician of the patient's response (test outcome) before proceeding to select the next most informative test. Fig. 4.3 shows a typical example of a 'print-out' from the terminal.

4.5 Results

The first achievment of the study came when both the off-line version of the program (in Algol 60 and Fortran IV) and the on-line version (G.E.I.S. and Cotan 3 versions) were finally run without error.

The results of the three experiments are shown in Figs 4.4 and 4.5. The overall success rate using the three sets of prior probabilities (.33, .33, .33), (.10, .89, .01), (.566, .316, .116) are shown in Fig.4.4.

The first general observation that can be made is that the higher the prior probability for a diagnosis, the fewer the misclassifications in that category in the 60 test cases; this is especially clear in Hashimoto's disease and thyroid carcinoma. The best overall model is that using equiprobability with the third model being close behind.

Misclassifications tend to occur in the direction of/

				Misdiagnosed as		
	Study	Prior	Correct	Wrong	H	S	C
	1	.33	32	2	0	0	2
HASHIMOTO'S	2	.10	29	5	0	4	1
	3	.566	32	2	0	0	2
	1	.33	17	2	1	0	1
SIMPLE	2	.89	18	1	0	0	1
	3	.316	16	3	2	0	1
	1	.33	7	0	0	0	0
CARCINOMA	2	.01	5	2	0	2	0
	3	.116	7	0	0	0	0

TOTAL	Study 1	56	4 = 93 per cent
	" 2	52	8 = 87 per cent
	" 3	55	5 = 92 per cent

FIG. 4.4. DIAGNOSTIC ACCURACY FOR 3 STUDIES.

of the other disease with the highest prior probabilities. For example, in Hashimoto's disease the misclassifications with the .10 model are mostly as simple goitre whose priors in this model are .89; a similar trend is found with misclassifications of carcinomas as simple goitre. The other misclassifications are probably too evenly spread to merit comment.

An important aspect of misclassifications is their relative seriousness clinically. Thus, only two misclassifications of carcinoma occur. Both of these are found where the prior for carcinoma is lowest of all (i.e. .01) and the misclassifications are both as simple goitre (where the priors are highest of all, i.e. .89). In both the equiprobability model and the third model there are no misclassifications of carcinoma at all.

The problem of misclassifications can be viewed in clinical terms by looking more closely at some examples of misclassifications. Thus, in the first model (.33, .33, .33) there were four misclassifications. Two simple goitres were misdiagnosed, one as Hashimoto's disease and one as carcinoma. Two cases of Hashimoto's disease were misdiagnosed as carcinoma.

In the case of simple goitre misdiagnosed as carcinoma/

			Number of Tests				
	Study	Prior	≤2	3-6	7-10	≥11	30
HASHIMOTO'S	1	.33	12	12	3	4	1
	2	.10	10	16	3	0	0
	3	.566	15	9	3	1	2
SIMPLE	1	.33	0	3	7	3	4
	2	.89	10	5	0	0	3
	3	.316	0	4	5	2	5
CARCINOMA	1	.33	0	5	1	1	0
	2	.01	0	1	2	2	0
	3	.116	0	2	2	2	1
					A		B

A . Cases with a final probability of ≥.99

B . Cases with a final probability of .51-.98 in which all 30 tests are used.

FIG. 4.5. NUMBER OF TESTS USED IN TEST CASES.

carcinoma, the thyroid gland was hard and nodular with a recent increase in size and, in fact, went to histology to exclude carcinoma. In the second case of simple goitre misdiagnosed as Hashimoto's disease, the precipitin and complement fixation tests were mildly positive even though histology showed a simple goitre . This case, of course, raises the very important question of dual pathology, i.e. the co-existence of simple goitre and a mild degree of auto-immune thyroiditis, which is well documented in the literature (Gribetz et al, 1954).

Review of the two cases of Hashimoto's disease misdiagnosed as carcinoma showed evidence strongly suggestive of carcinoma. Both patients were euthyroid, both goitres were hard and no immunological evidence of Hashimoto's disease was demonstrated despite the clear histological diagnosis. As was stated earlier, the final diagnosis was established in each case either histologically or by agreement among clinicians well experienced in thyroid disease.

Another very important aspect of the study is summarised in Fig. 4.5. This shows an analysis of the cases which were correctly diagnosed in terms of the number of tests/

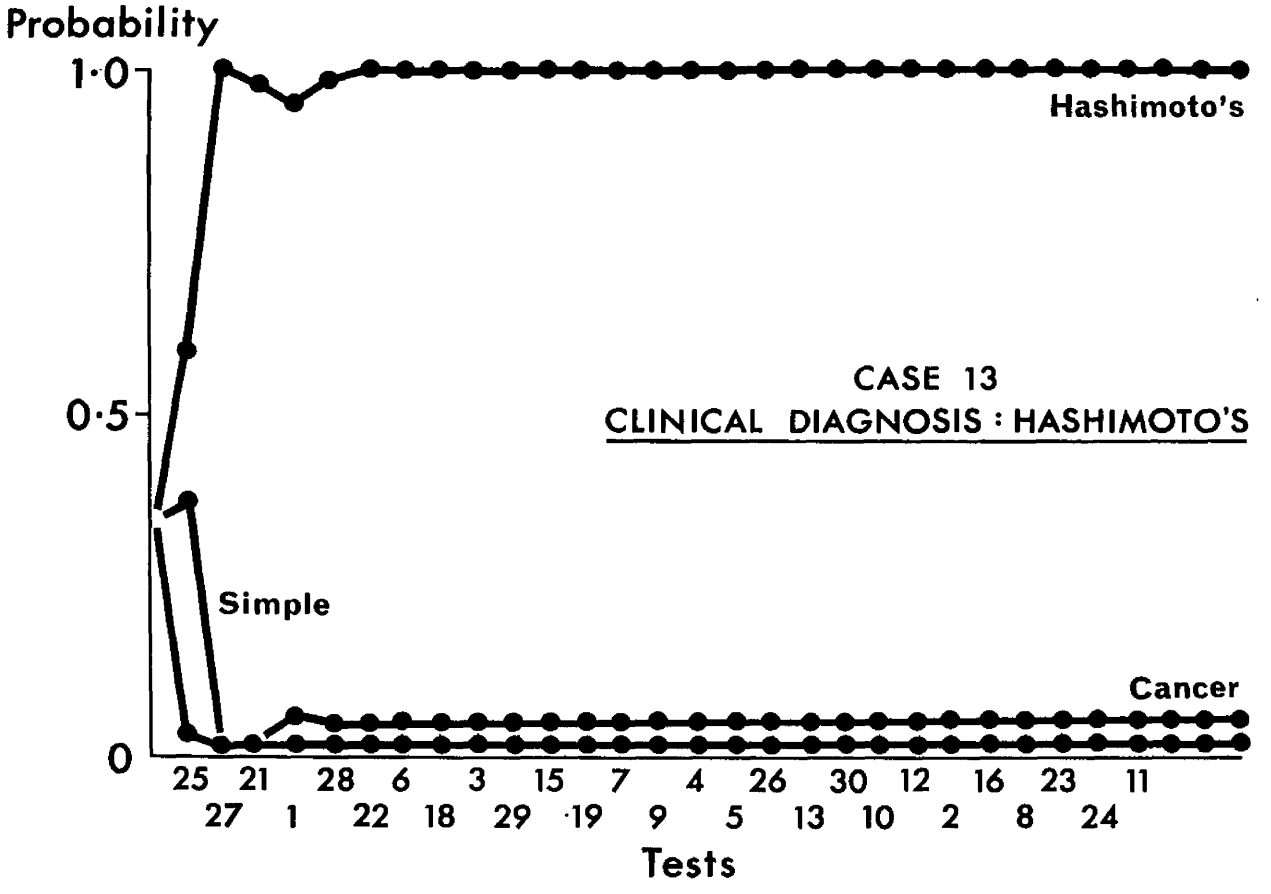
tests needed to reach and stay at a probability of 0.99 and over; also shown are the cases where the diagnoses were correct but where the final diagnosis was less than 0.99.

Particularly striking in these results is the fact that over a third of all the Hashimoto's cases were diagnosed with only two tests or less in the first and third experiments; in all three experiments over two-thirds of all cases took less than 7 tests. This effect was least marked in the carcinoma cases. The effect of the high prior probabilities in reducing the number of tests is particularly marked in the second experiment with simple goitres.

In each case in the first experiment the most informative test was the consistency of the gland (i.e. in all 60 test cases). In, for example, the 12 cases of Hashimoto's disease in the first experiment, the second 'best' test were the complement fixation test (10), the PBI^{127} (1) and the precipitin test (1). The most informative test in both the second and the third studies was the complement fixation test.

4.6 Illustrative cases

The value of a sequential approach is best seen when the/



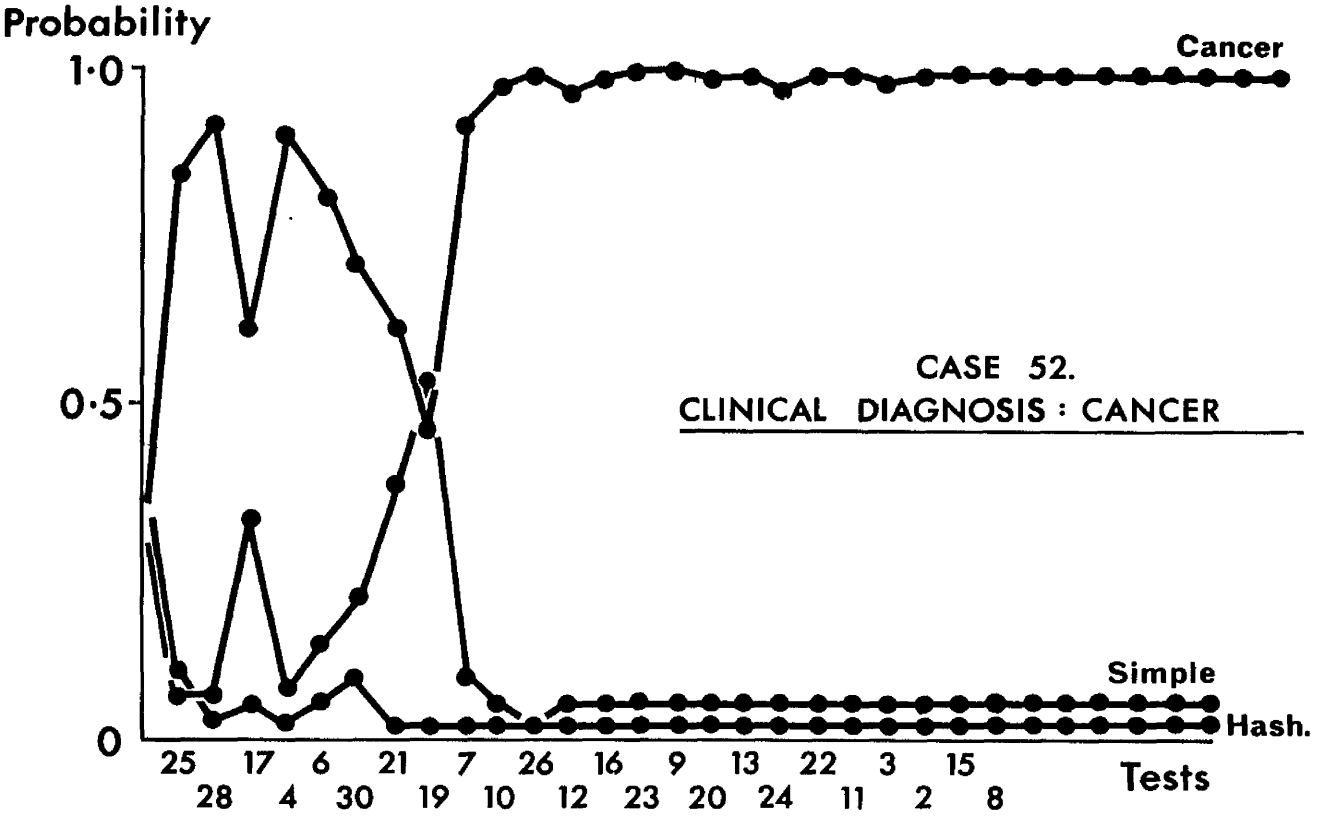
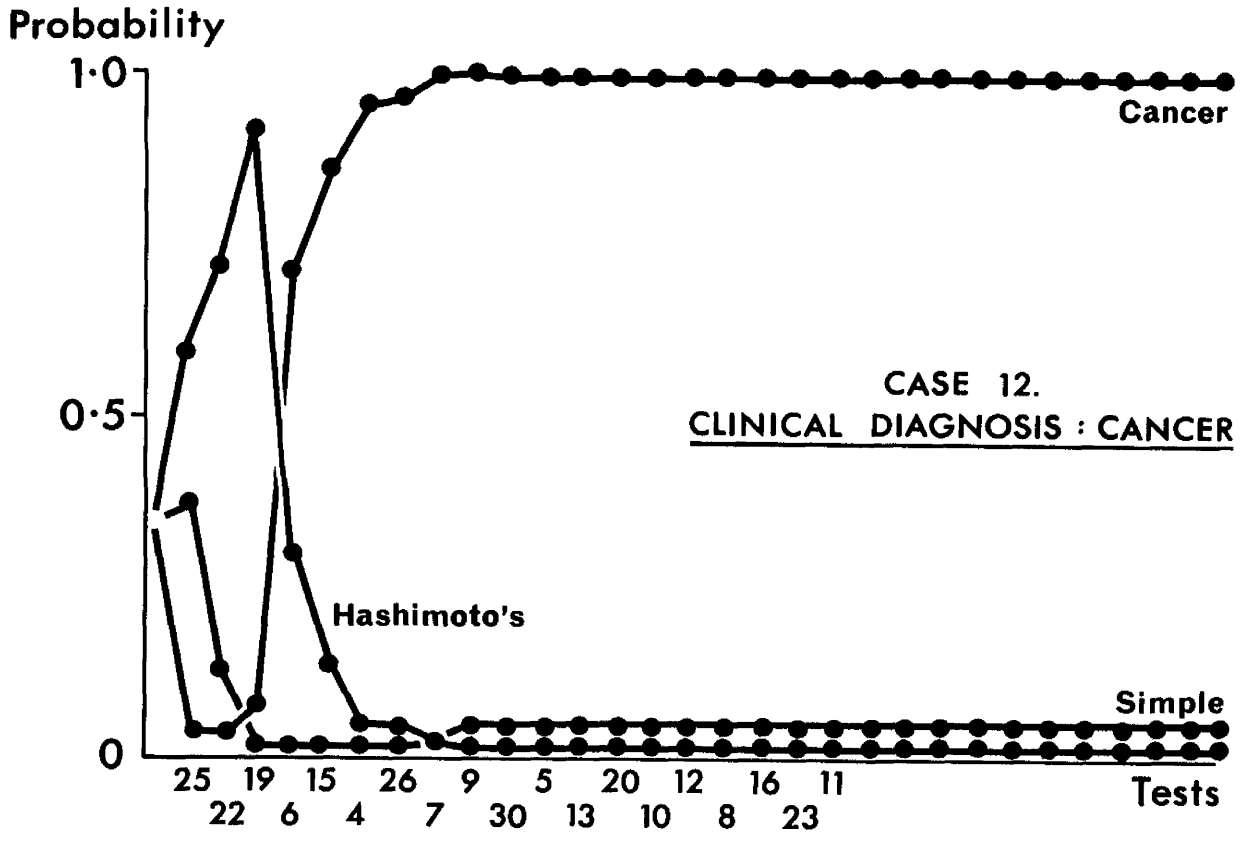


FIG. 4.6. GRAPH FOR CASE TAKING LESS THAN 7 TESTS.

FIG. 4.7. EXAMPLE OF EFFECT OF CONTRADICTORY INFORMATION.



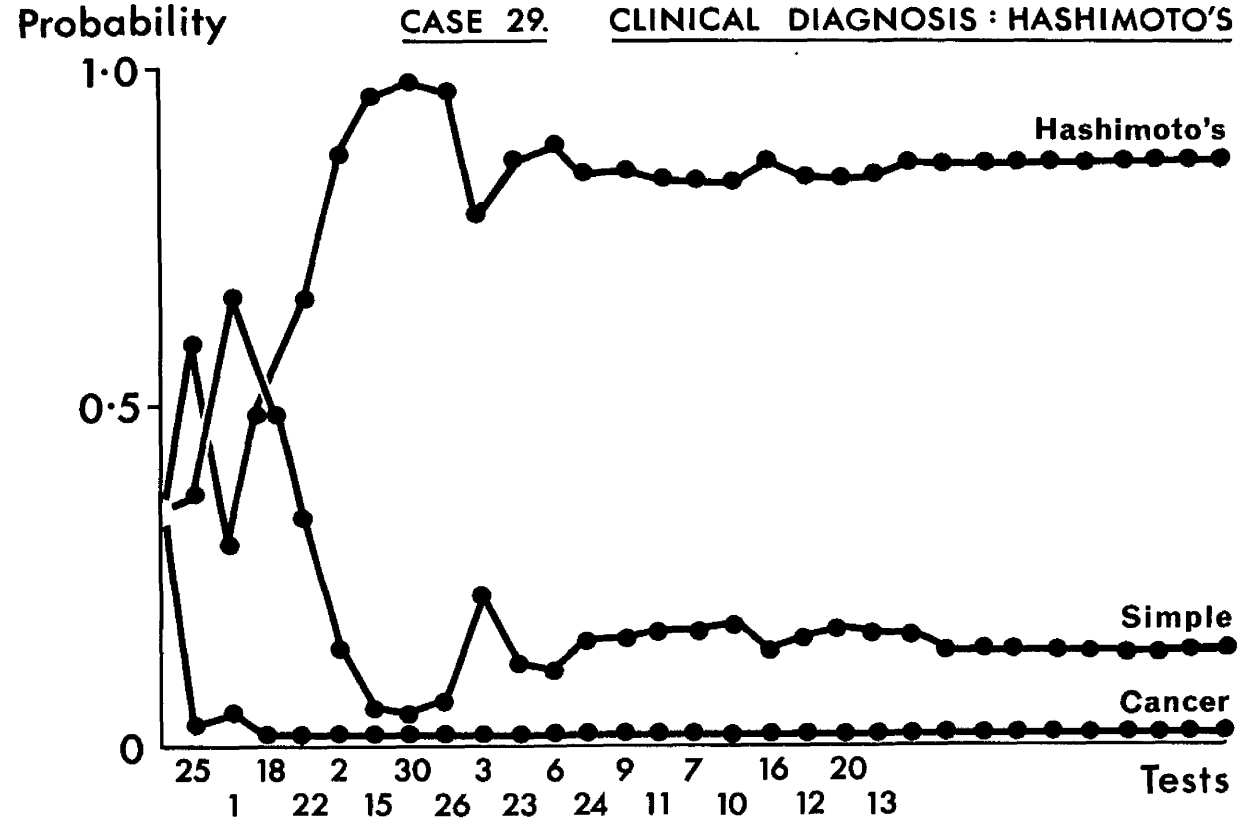


FIG. 4.8. TYPICAL GRAPHS OF TEST CASES

FIG. 4.9.

the changes in probabilities (as tests are selected) are displayed graphically. Fig. 4.6 shows a typical case which took less than seven tests to reach a probability of 0.99. Fig. 4.7 shows a case where contradictory information led to the probabilities oscillating in the early stages before settling later to a final level of 0.98. The outcomes of tests 4, 6, 30, 21, 19, 7 and 10 were:

4.	Tracheal deviation	Yes
6.	Fixation to tissues	No
30.	Age	31-60 years
21.	PBI^{127}	$<3.0\ \mu G/100ml$
19.	E.S.R.	0-20 mm.Hg in 1st hour
7.	Cervical lymph glands	Palpable
10.	Hoarseness	Yes

The sequential technique allows us to pinpoint the poor or contradictory items of evidence (tests 4, 6, 30, 21, 19) and demonstrates how the positive evidence of tests 7 and 10 overcomes this. Figs. 4.8 and 4.9 show similar graphs of other cases.

4.7 Discussion

The place of this model in computer-assisted diagnosis can be discussed under a number of headings.

4.7.1/

4.7.1 Sequential v 'static' model

In the introduction to this thesis the sequential nature of clinical decision-making has been discussed (Fig.1.2). Almost all studies in the field of computer-assisted diagnosis have used the 'static' diagnostic model (e.g. Warner et al, 1961; Lodwick et al, 1965; Fitzgerald et al, 1966). The exceptions are the logical decision trees. Thus, Kleinmuntz (1965) used a simple "Yes/No" type of decision tree in the diagnosis of neurological disorders which closely resembles that of Wortman (1966) while Edwards (1970) has developed a similar decision tree for the diagnosis of dysphagia. Among the commonest uses of logical decision trees are the computer based history-taking systems using a computer terminal, of which those of Slack et al (1966) and Coltart et al (1965) are the best known examples. History-taking systems of this type are briefly reviewed in Taylor (1970c).

All of these systems must of necessity simplify the complex interaction and exchange of information between clinician and patient and tend to be rigid. Clinical situations where a purely logical exchange of information is possible are very limited. An incorrect answer to a question/

question will result in a wrong and irreversible path being selected down the decision tree.

The probabilistic sequential model used in this study approximates much more closely to reality (as studies of diagnostic skill in Chapter 5 will show) and is much more flexible. It is also able to cope with misleading information as the analysis of the illustrative cases (4.6) has shcwn. It is of wide application in clinical medicine and is likely to make the use of computer based history-taking much more feasible. Many attempts have been made to use long questionnaires to formalise history-taking, for example the Cornell Medical Index-Health Questionnaire (Erdmann, 1964) and the Minnesota Multi-phasic Personality Investory (Rome, 1962). The latter consists of 550 Yes/No statements. The use of a probab-ilistic decision tree might greatly reduce the number of items needed for a diagnosis to be made.

4.7.2 Costs

The model described here concentrates solely on the information processing aspect of the diagnostic process and takes no account of the costs of the tests or the very important problem of non-financial costs (such as the discomfort/

discomfort and inconvenience of tests and the 'cost' of misdiagnosis). The role of statistical decision theory in providing a framework for formulating such cost considerations has been discussed earlier (2.4).

The only non-logical decision tree so far developed, apart from the one described here, is that of Gorry and Barnet (1968). These authors developed a theoretical decision tree in which arbitrary costs were attached to the outcome of clinical decisions. The choice of decision was decided by computing the expected costs at each decision point and by selecting the choice with the lowest expected costs. The cost values in this study were not derived from clinicians themselves or from the study of their behaviour.

Aitchison (1970) has put forward a mathematical technique for inferring the costs which clinicians attach to the treatment which they select. The author is at present engaged in applying this technique to the study of treatment selection in thyrotoxicosis.

A less complex approach to costs would be to concentrate solely on financial costs and by attaching a cost to/

to each test, develop the basic computer program to select the test with the highest expected yield of information per unit cost. A cost analysis has been completed in the Thyroid Clinic at the Royal Infirmary, Glasgow, and a modification to the computer program to take account of costs is being made. It is quite clear from the study in the following chapter of the cost estimates made by clinicians that true financial costs of investigations are not accurately known to them.

4.7.3 The effect of prior probabilities

The effect of the prior probabilities are summarised in Figs. 4.4 and 4.5.

The first general observation is that the first model (.33, .33, .33) is the best from the point of view of avoiding misclassifications. The second model had the highest total misclassifications and also wrongly diagnosed two cases of carcinoma. In a model based on 'costs' the possible cost of misdiagnoses such as these would have to be taken into account. Boyle and Anderson (1968) made some attempt to guard against such misdiagnoses in effect by attaching weights to each diagnosis so that such misclassifications, as missing a case of thyroid carcinoma, are made much less likely. The same authors later/

later compared two models in diagnosing non-toxic goitre. These were the normal Bayesian' conditional probability model, and a modification of this, in which the prior probabilities were not taken into account (relative likelihood).

The authors in this latter study were in effect comparing model 2 (.10, .89, .01) the full Bayesian model, with (.33, .33, .33) the relative likelihood model. They also reached the same conclusion that the latter model was more effective in that it produced fewer misclassifications and that no cases of carcinoma were misdiagnosed. The same model effectively weights the diagnosis in favour of carcinoma (.33) compared with the actual prior (.01).

The overall effect, however, of the prior probabilities seems to be small. This conclusion is of some importance since the difficulty of matching the prior probabilities to different population samples (e.g. G.P.surgery, general medical clinic, thyroid clinic) is then avoided.

4.7.4 Redundancy of clinical information

There is almost certainly a great deal of redundancy in data recorded on patients.

Fig. 4.6 for example, demonstrates clearly that the great/

great majority of the tests in this case were superfluous. Fig. 4.1 shows several tests under the heading 'serum tests'; of these, 1, 28 and 27 are much more discriminating than the others (as shown by the fact that they are selected much earlier in the diagnostic sequence). Many tests could be eliminated without loss of diagnostic accuracy; for example, test 8 (pyramidal lobe) is almost invariably well down the list of the tests selected and this is explained by the likelihoods for this test (Fig. 2.1) which show that its rate of occurrence is almost identical in each disease. It should be possible to eliminate tests by noting their relative position in the test selection in a large number of test cases and removing those which are lowest on the list.

The variation among tests in their discriminating power is also revealed by the graphs (Figs. 4.6 to 4.9) and a poor test (e.g. pyramidal lobe) can be precisely located using this approach. Alteration of the 'normal' range in a laboratory test in a particular group of diseases may make a test more precise. Thus, Boyle et al (1966) adopted an informal method of classifying continuous variables to achieve best separation in the group of diseases being studied. The classification of the/

the erythrocyte sedimentation rate into three classes (0-20, 21-40, > 40 mm in 1st hour) was decided after plotting a frequency histogram and noting that the above classification achieved the best arbitrary separation in cases of non-toxic goitre. The important point to note here is that a test is being altered for operational reasons, i.e. to be of maximum value in diagnosing a group of patients, rather than to make a more precise measure of a physiological variable.

The discriminating power of clinical data (as opposed to laboratory or other instrumental data) may well be greatly underestimated. Thus, the test invariably selected as best in all 60 cases in model 1 (.33, .33, .33) is the 'consistency' of the thyroid gland. This is classified as 'hard', 'firm' or 'soft'. It would appear that the inevitable observer-error in such a sign has not diminished too greatly its discriminating power in this group of diseases; in the second and third experiments (.10, .89, .01 and .566, .316, .116) this test almost invariably ranks second. The implications of this feature are considerable, since this clinical sign is therefore of comparable diagnostic value in this group of diseases to such elaborate and expensive laboratory tests as 24-hour thyroidal ^{131}I uptake and the 48-hour $PB^{131}I$.

The influence of the prior probabilities on the redundancy of tests is not entirely clear (Fig. 4.5). Thus, if one uses as a guide the number of cases where less than 11 tests were needed, then the second experiment with 4 7 out of 60 is better than the first (43/60) or the third (40/60). However, account must be taken in deciding on the overall efficiency of the misclassifications already discussed. Therefore, the difference between the second and the first is not high enough to deflect one from the choice of the first model as the overall best.

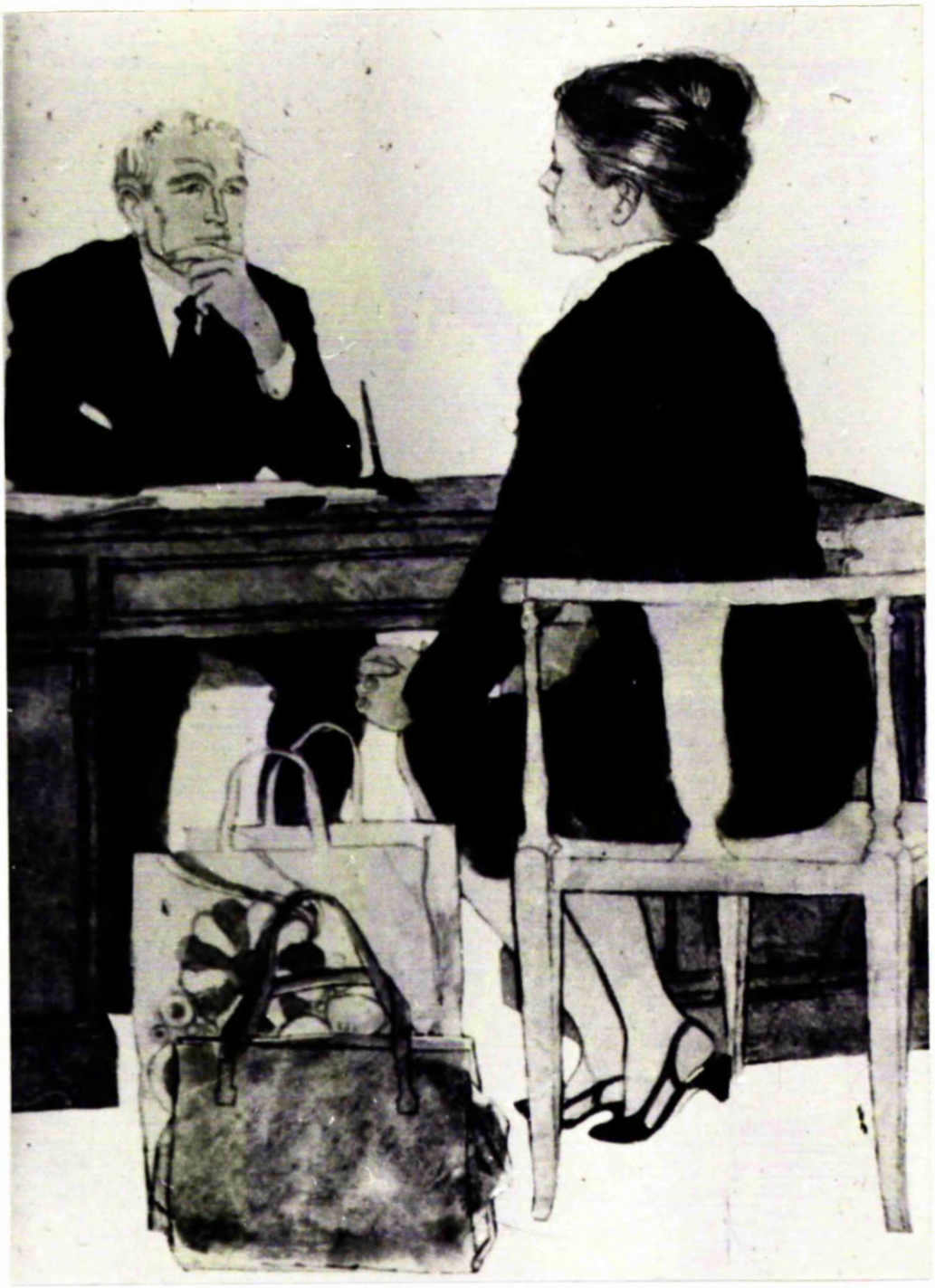
Clearly, the complex balancing of all such factors enters into clinical decision-making and by giving clinicians fresh insight into the elements of the diagnostic process we hope to improve diagnostic performance (Aitchison, 1970a, 1970b).

The final choice of a particular model in any given clinical situation must be reached by weighing up such factors as the effect of the priors selected, the arbitrary level for a final diagnosis (the level of .99 in this case represents an error rate of 1 in 100 cases), the factor of redundancy of tests, the choice of 'normal ranges' and methods of classification of continuous data and, above all, the importance of financial and non-financial costs.

4.7.5 Programming and system design

This study has shown that the use of an on-line computer terminal with suitable programs makes the use of a computer in diagnosis at least technically feasible in an outpatient clinic. In the epilogue of this thesis dealing with future studies a plan is put forward for a computer-based screening clinic, which is under development at the Department of Medicine in the Royal Infirmary, Glasgow.

It is important to note that this model of non-toxic goitre involves only 30 tests in three diseases. The time delay between entering a patient's response (test outcome) at the terminal and printing out the next 'best test' is of the order of 5.5 - 6.0 seconds. In a larger model of all thyroid diseases (with 10 or more diseases and 80 tests) the time delay may be much longer. However, the computer used in the present study was relatively small and it is likely that the larger time sharing systems now under development could successfully manage such large calculation fast enough to make on-line diagnosis practical.



CLINICAL DECISION MAKING

C H A P T E R 5 : DOCTORS AS DECISION-MAKERS

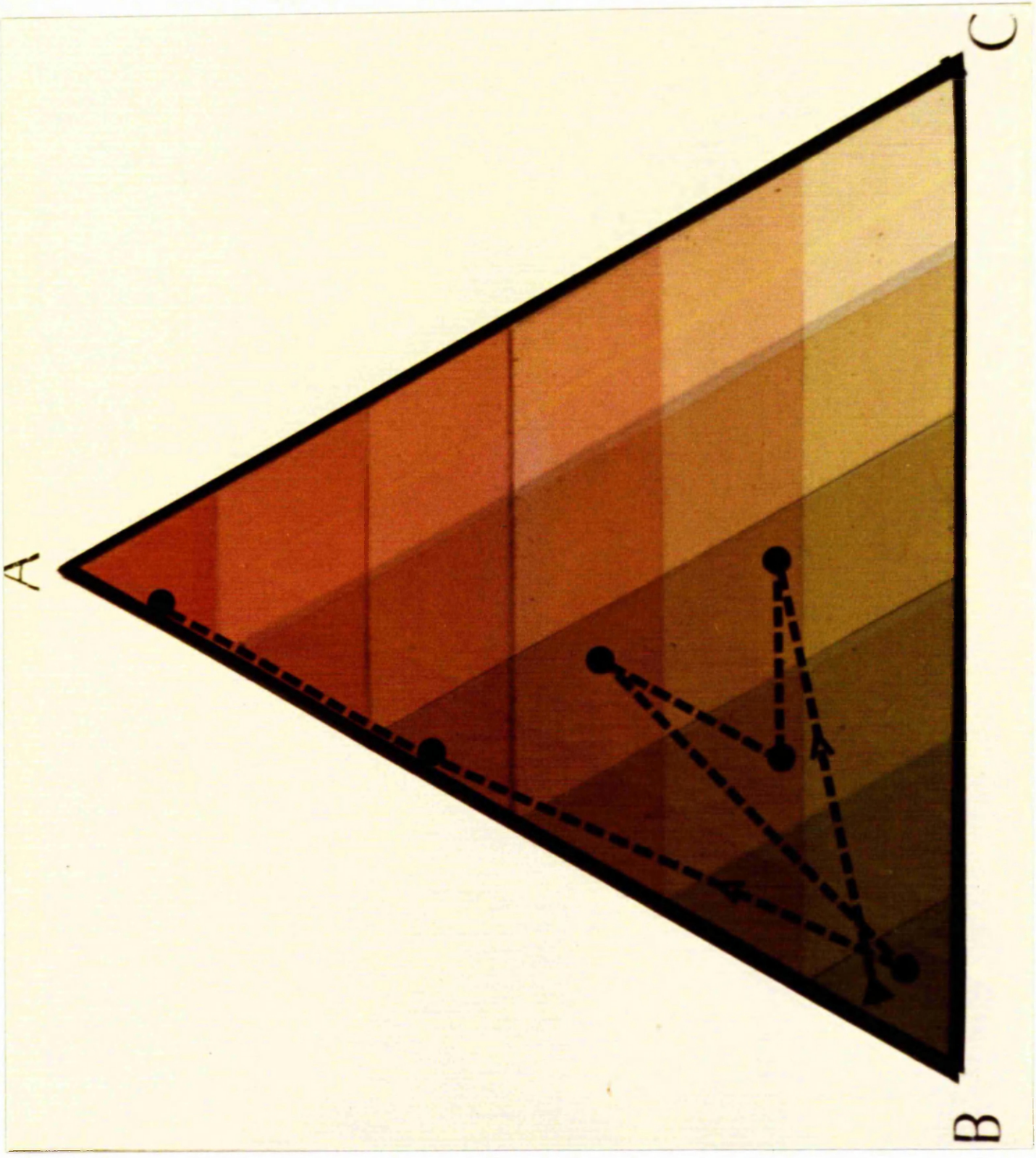


FIG. 5.0. COLOURED TRIANGLE WITH TYPICAL PATH OF CLINICIAN TO DIAGNOSIS.

5.0 Summary

In a study of six experienced clinicians tackling 20 varied cases of non-toxic goitre, the clinicians were compared in detail with a computer program whose accuracy in diagnosis was comparable to an experienced clinician.

The techniques of information theory, statistical inference and conditional probability theory were used to produce a diagnostic profile of each clinician for each of the 20 cases which he tackled. This profile, consisting of five measures of discrepancy between each clinician and the computer program, allowed the six clinicians to be compared indirectly with one another.

Using the 'entropy' calculation from information theory it was possible to demonstrate, by means of equilateral triangles, the great variability in diagnostic strategies among the group of clinicians.

Measures of personality including extroversion and obsessionalism showed significant correlation with individual elements of the diagnostic profile, with diagnostic accuracy and with the total number of tests used to diagnose the 20 cases in the study.

General/

General clinical experience, but not specialist thyroid experience, appears to be related to other elements of the profile. The possible relevance of these findings to the teaching of the diagnostic process and the selection of sub-specialities by clinicians is discussed.

It is observed that individual clinicians may vary in their capacity for processing information and the relevance of this finding to clinical decision-making is noted.

The possibility that two measures in the diagnostic profile, liberalism and 'diagnosis', may be used to detect a type of pattern recognition is advanced. Results which may indicate that one of the six clinicians in particular uses this type of information processing are presented.

Finally, the role of financial and non-financial costs in clinical decision-making is discussed in the light of estimates by the clinicians of the financial costs of the laboratory tests used in the study.

5.1 Introduction

The diagnostic process can be viewed as a sequence of decisions and is then amenable to analysis in terms of information theory, probability theory and statistical decision theory.

The/

The sequential model presented in the last chapter has a diagnostic accuracy of 93 per cent (Fig. 4.4) and so is comparable in this small set of diseases to that of an experienced endocrinologist.

If clinicians are to accept the value of such examples of computer-assisted diagnosis, a comparison between the computer program and the diagnostic behaviour of the clinician is necessary.

Such comparisons in the past have been based on simple differences between the diagnostic accuracy of the clinician and that of the computer. Several authors have, for example, used differences in percentages of correct diagnosis or in the average probability ratings (Crooks et al, 1959; Warner et al, 1961; Boyle et al, 1966). Moreover, in such studies the diagnosis has only been made after the process of data collection is completed (the 'static' or traditional model of the diagnostic process (Fig. 1.1)).

The sequential process of diagnosis (Fig. 1.2) can by analysed further into a simple single recurrent cycle (Fig. 5.1). At each such cycle the following factors or 'costs' must be weighed up:

(i) the advantages of making an immediate correct diagnosis,

(ii)/

STEP	DECISION
1	Select a test
2	Carry out test and observe outcome
3	(i) Select further test, i.e. return to Step 1 or (ii) make diagnosis.

FIG. 5.1. SINGLE RECURRENT CYCLE.

(ii) the consequences of making an over-hasty misdiagnosis

(iii) the financial cost, discomfort, inconvenience and delay due to further diagnostic investigations before treatment is begun.

The role of financial costs in diagnostic decision-making within the National Health Service is far from clear. It would obviously differ from that in other countries where clinicians are of necessity more cost conscious.

In the design of the study an attempt has been made to eliminate such cost-consciousness as may exist by constantly reminding the participants during each case that tests were to be regarded as equally available and 'cost-free'.

The primary purpose of the study is to show how a detailed comparison between computer and clinician is possible. Objective techniques have been developed which allow the diagnostic performances of a group of clinicians to be compared with each other in some detail. The simple three disease system of non-toxic goitre is used and six clinicians are studied in detail. The influence of personality factors and of clinical experience is also shown.

5.2/

	TESTS	1st SESSION	2nd SESSION
A	≤ 2	2	5
B	3-6	3	2
C	7-10	1	0
D	≥ 11	1	0
E	$<.99$	2	2
F	Wrong	1	1

GROUP A,B,C, D $\geq .99$

GROUP E $< .99$

H = 2, S = 16, C = 2

1st Session: 2, 9, 28, 12, 13, 20, 17, 11, 4, 23

2nd " : 36, 1, 2, 23, 11, 15, 24, 25, 17, 16

FIG. 5:2. EXPERIMENTAL DESIGN FOR STUDY.

5.2 The study

A total of 20 cases of non-toxic goitre was used, comprising 16 different cases with 4 of these repeated unknown to the participants. The cases were presented to each of six clinicians, members of staff in the Thyroid Clinic of Glasgow Royal Infirmary, in 2 sessions of 10 cases. The design of the study is shown in Fig. 5.2. The cases consisted of 2 cases of Hashimoto's disease, 12 cases (4 repeated) of simple goitre and 2 cases of thyroid carcinoma; this distribution was chosen as reasonably representative of the frequencies (10 per cent, 89 per cent and 1 per cent) with which these diseases are seen in the clinic.

The cases were selected to provide variation in difficulty by the objective method of classification used in the previous chapter (Fig. 4.5). The method groups cases according to the number of tests required to reach a diagnostic probability level of 0.99 by the procedure used in the program (i.e. the combination of Bayes' theorem and the 'minimal entropy' calculation (2.2)). The cases were presented to each clinician in an identical random order (Fig. 5.2) such that no clear pattern of diagnosis or ease of diagnosis was apparent.

5.3/

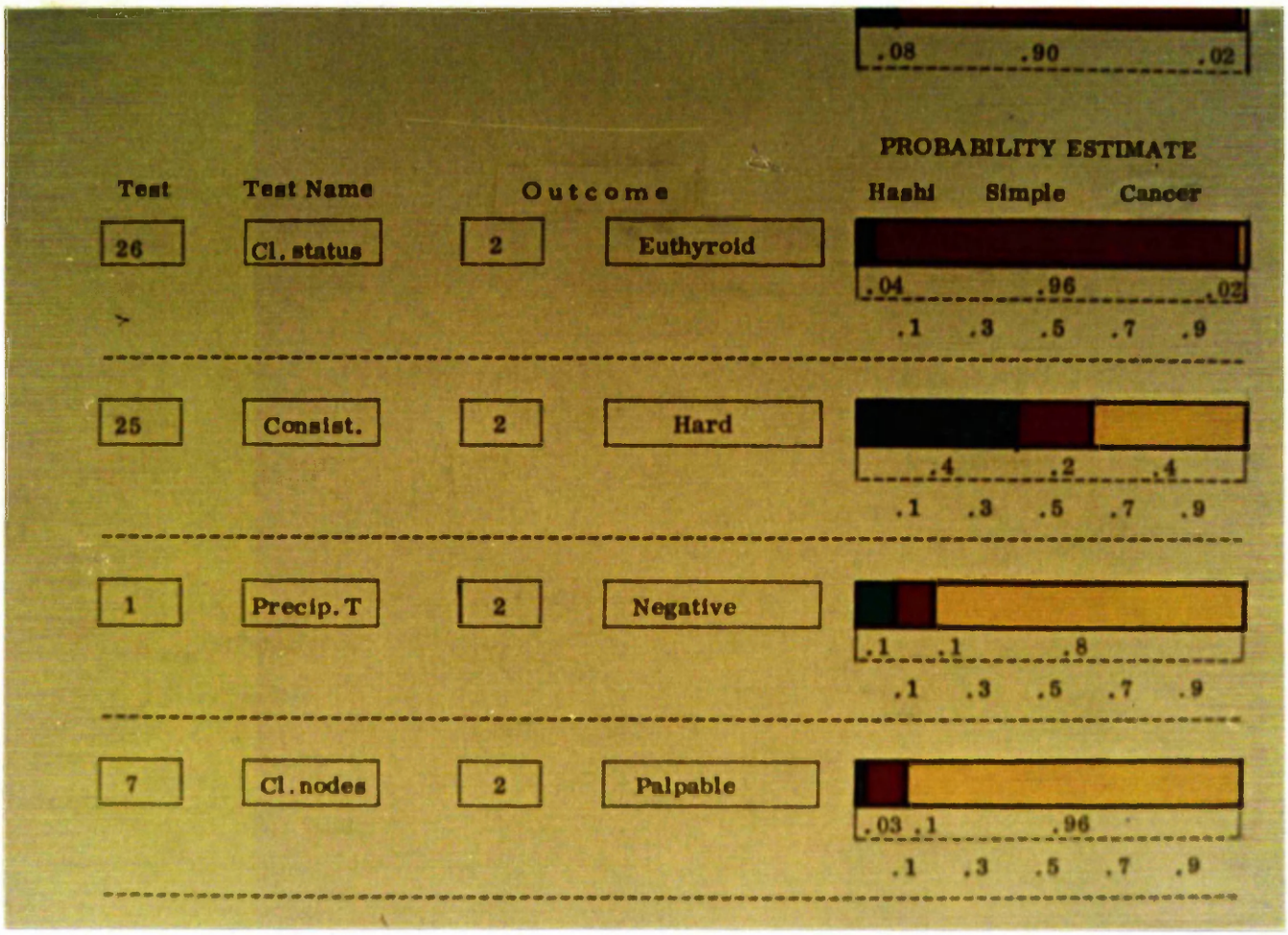


FIG. 5.3. RECORDING SHEET USED BY CLINICIANS.

5.3 Technique

The cases were presented in the form of abstracted case records with the results of 30 tests (Fig. 4.1) available on each. The clinician used a recording sheet (Fig. 5.3). He began by recording his assessment of the prior probabilities (incidence rates) of the three diseases at the clinic. He then selected his first test, was told the outcome and then entered his revised assessment of the probabilities of the three diseases in the light of this information. He noted this by dividing the strip (of unit length in Fig. 5.3) into three proportions, and also recorded the probabilities below. He then selected his next test, was told the outcome and again altered the probabilities as he thought fit. He continued in this way until he was satisfied that he had sufficient information to make a diagnosis.

He was allowed to select tests in any sequence, to 'back track', for example from a laboratory test to an item of history. He was repeatedly reminded that all tests were equally available and that no costs should be taken into account.

5.4 Analysis

The principles underlying the analysis depend on statistical/

statistical ideas of inference and decision theory, and of information theory, particularly the work of Shannon (1948), Kullback and Leibler (1951) and Lindley (1956).

The computer program is that already described in the previous chapter which operates on the following assumptions:

(i) for a given disease, the tests are statistically independent;

(ii) the tests are equally available without cost;

(iii) the test selected at each diagnostic cycle (Fig.5.1) is that which, relative to the position attained, promises to provide most additional information by the end of the cycle;

(iv) after a test outcome becomes available, the probabilities of the disease are updated by the use of Bayes' theorem.

A clinician's decision-making behaviour at each diagnostic cycle is described by:

(i) his choice of test;

(ii) his updating of the disease probabilities after learning the outcome of the test.

At the start of each cycle the clinician is uncertain about the diagnosis. The degree of his uncertainty is indicated by/

by the current probabilities he is quoting and can indeed be quantified by means of the 'entropy' calculation (Shannon, 1948) described in Chapter 2.2 and used in the computer program described above. For example, the probabilities $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$ assigned to Hashimoto's disease, simple goitre and thyroid carcinoma correspond to a maximum degree of uncertainty since the clinician is equally 'torn' between the three possible diagnoses. At the opposite extreme, in such a probability assessment as (1, 0, 0), the clinician is stating that the diagnosis is certainly Hashimoto's disease and there is no uncertainty in this view. Such an assessment as (0.5, 0.3, 0.2) clearly lies intermediate to these two extremes in its associated degree of uncertainty. This intuitive ordering is reflected in the quantified degrees of uncertainty U calculated as 1.58, 0 and 1.48 bits (standard units for measuring uncertainty and information; see Shannon, 1948) for the probability assignments $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$, (1, 0, 0) and (0.5, 0.3, 0.2), respectively.

5.5 The equilateral triangle

An illuminating method of presenting a clinician's path to a diagnosis is to use an equilateral triangle

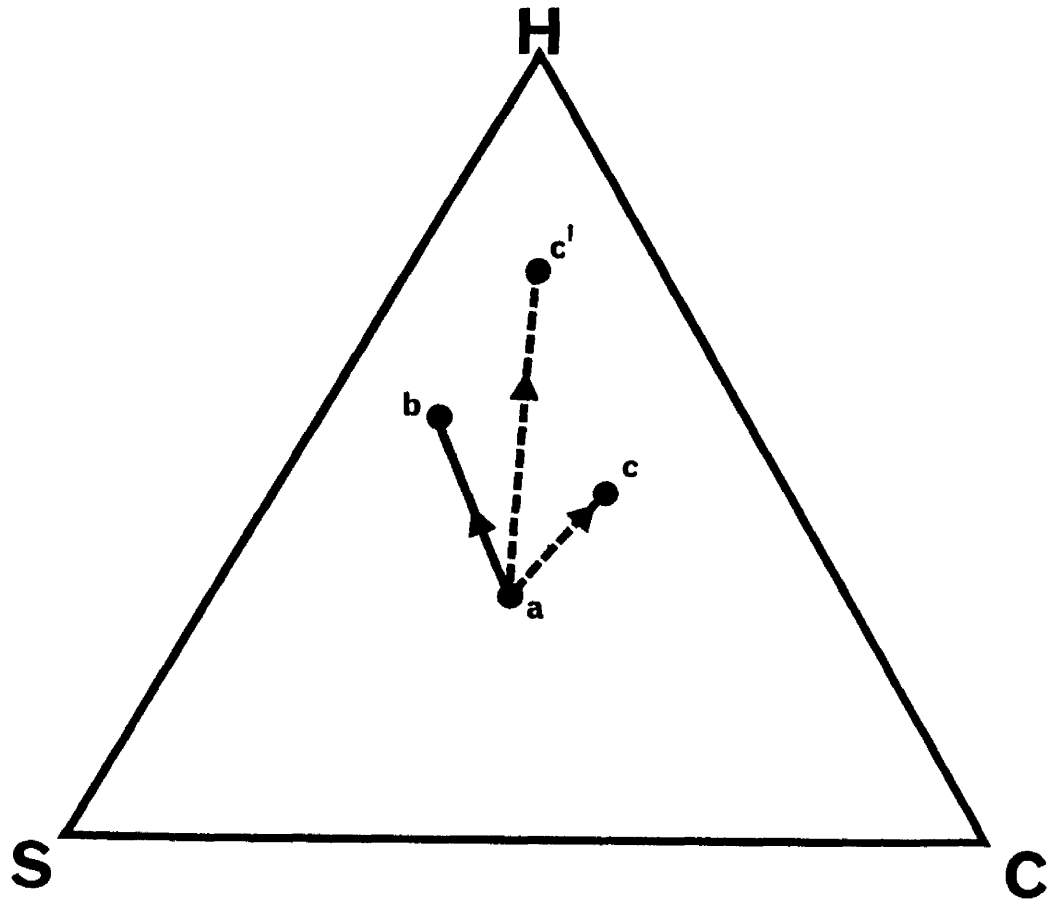


FIG. 5.4. EQUILATERAL TRIANGLE SHOWING CONSERVATIVE AND LIBERAL USE OF INFORMATION.

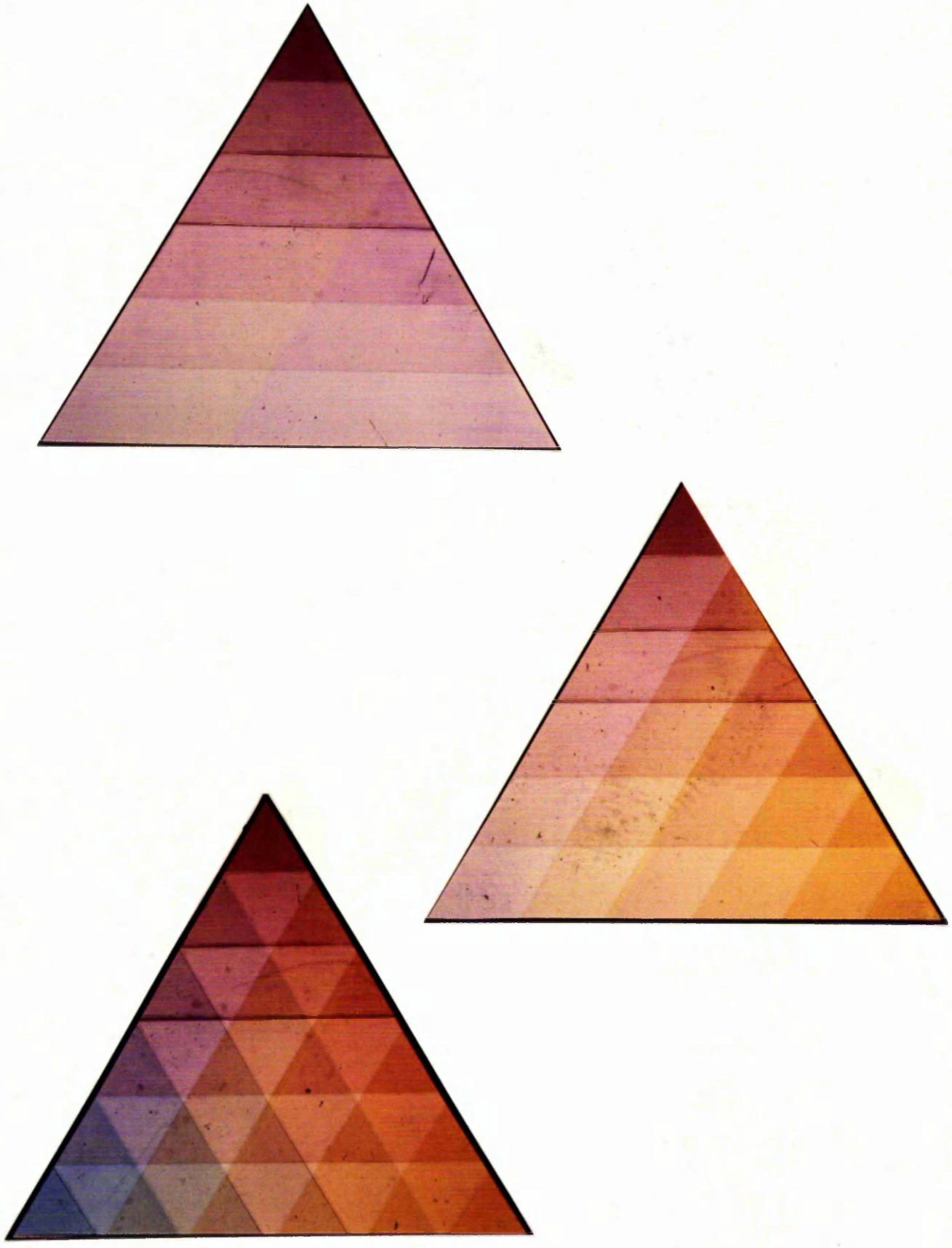


FIG. 5.5. PROBABILITY VALUES FOR EACH DIAGNOSIS REPRESENTED BY A COLOUR INTENSITY.

HSC (Fig. 5.4), whose vertices represent the three diseases: Hashimoto's disease, simple goitre and thyroid carcinoma. If the altitude of this triangle is of unit length then each point within the triangle can be used to represent a probability assessment for the three diseases. In this representation the distance of the point from the side opposite a disease vertex is the probability placed on that disease; for example, the distances of a in Fig. 5.4 from the sides SC, CH and HS are 0.5, 0.3, 0.2, and hence a represents a probability assessment of 0.5, 0.3, 0.2 on Hashimoto's (H), simple goitre (S) and thyroid carcinoma (C). Each time the clinician changes his probability assessment he moves from one point to another and so traces out a diagnostic path within the triangle (Figs. 5.5 and 5.6); his move towards a corner (for example H) represents his changing proximity to that diagnosis (Hashimoto's disease).

We can provide an extra dimension to this visual picture by associating with each point in the triangle its degree of uncertainty discussed above. If each degree of uncertainty is represented by a depth below the surface of the triangle we can trace out as the uncertainty surface a kind of triangular bowl (Figs. 5.6 and 5.7). The surface reaches/

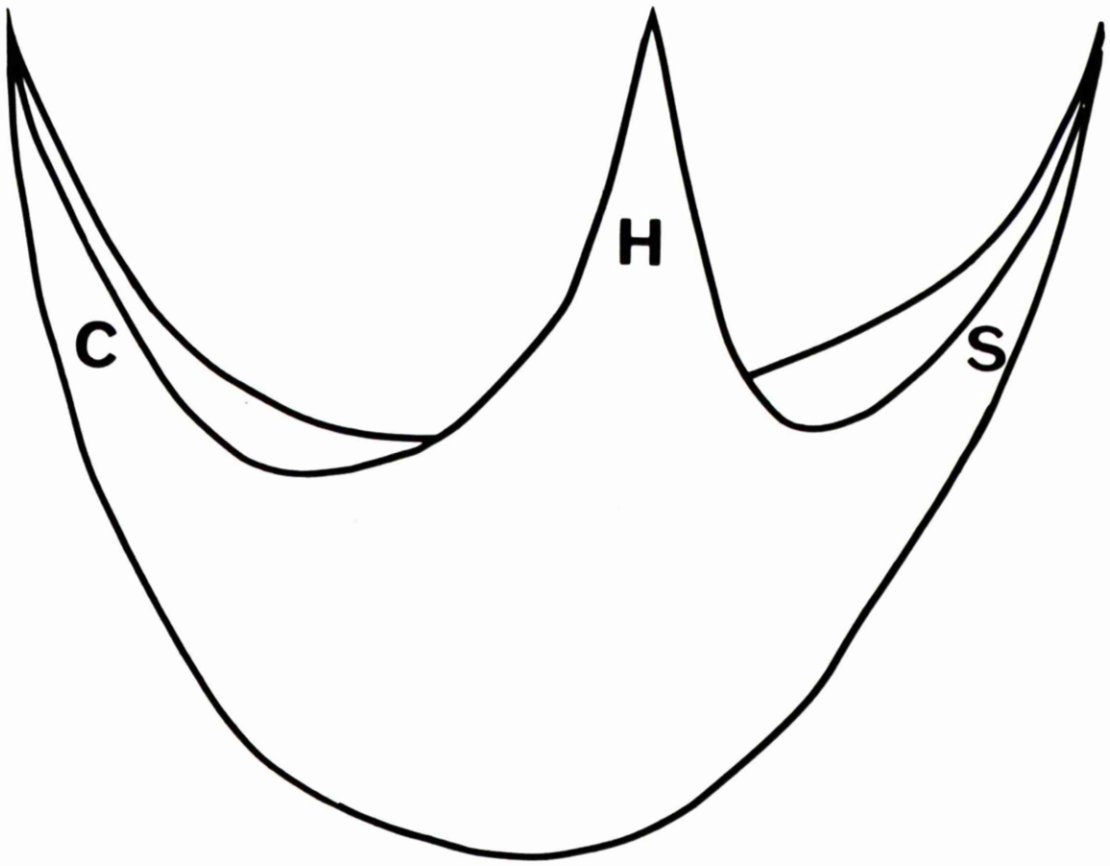


FIG. 5.6. TRIANGULAR BOWL.

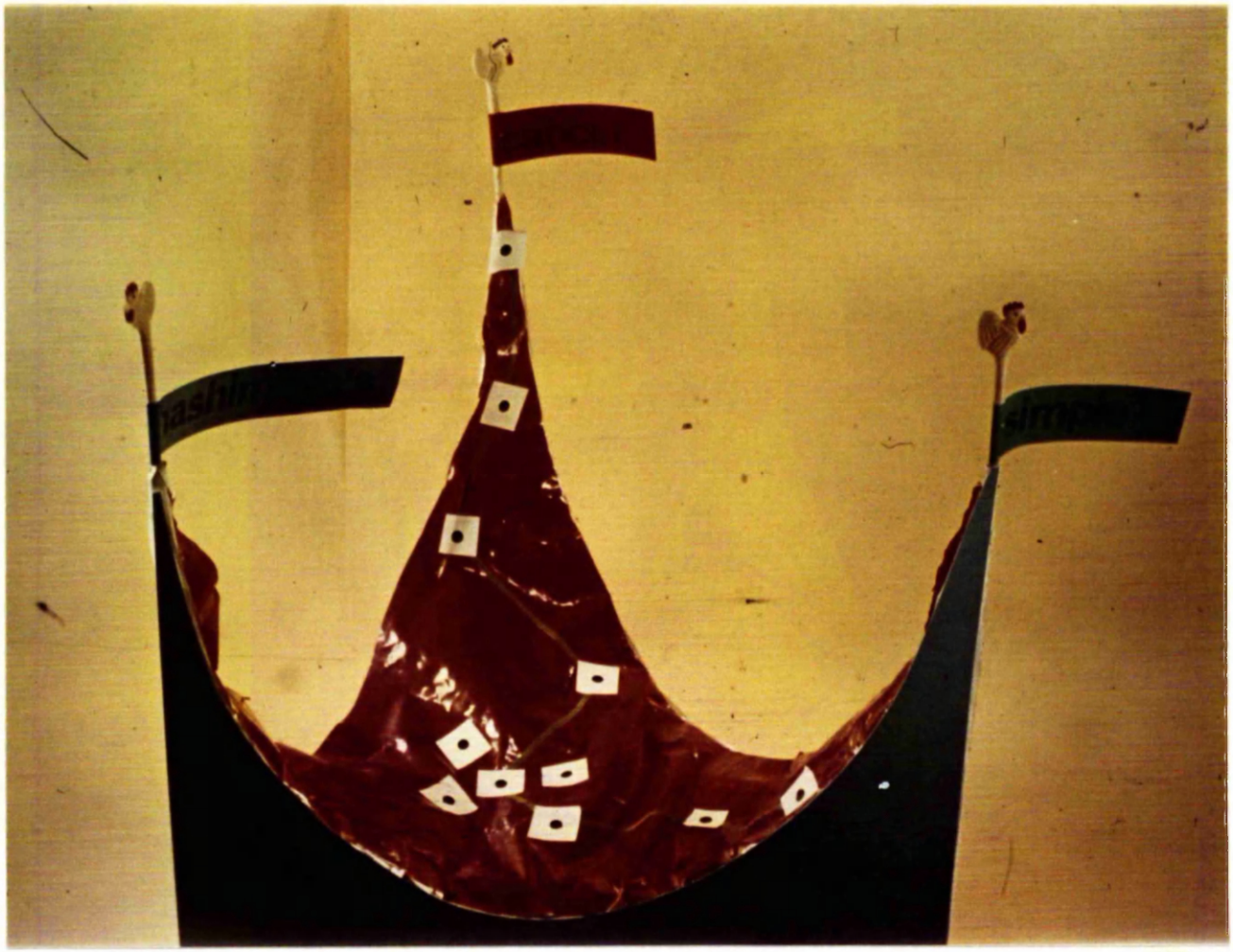


FIG. 5.7. MODEL OF TRIANGULAR BOWL WITH TYPICAL PATH OF A CLINICIAN.

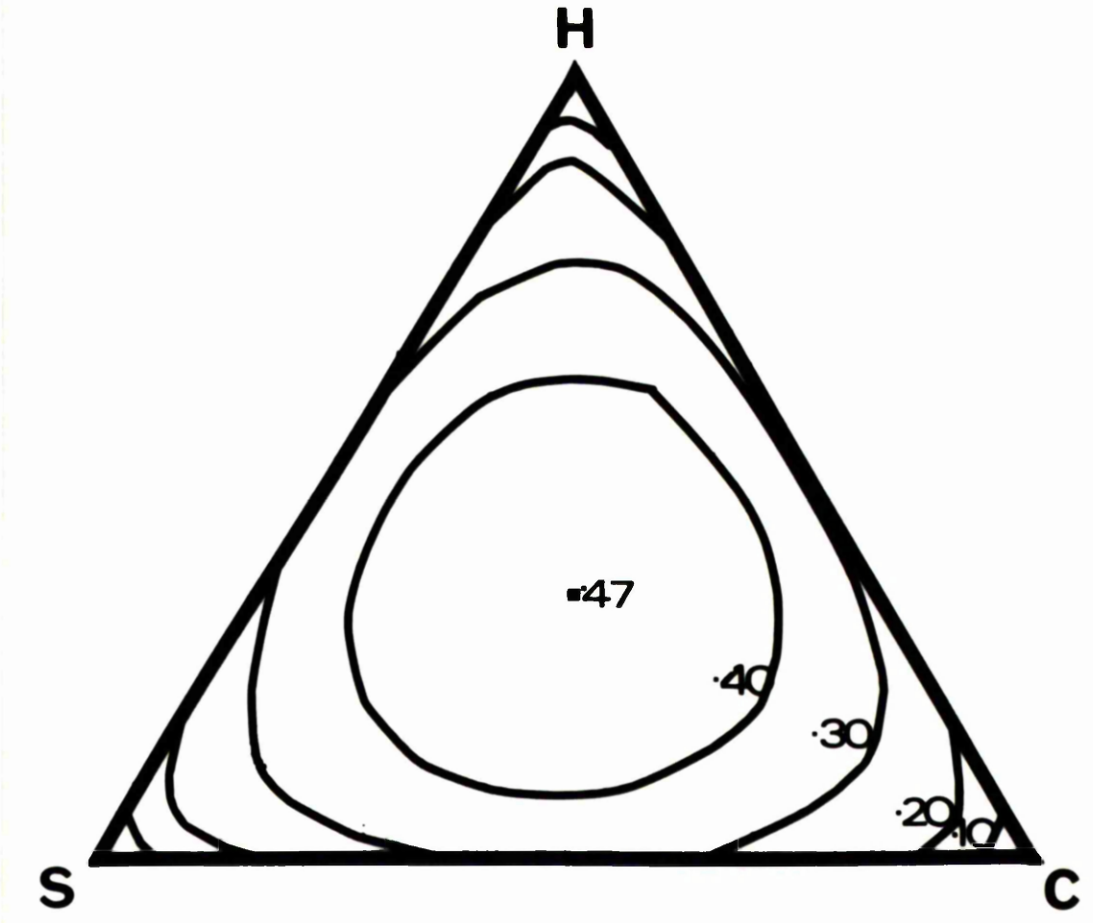


FIG. 5.8. TRIANGLE WITH CONTOURS OF UNCERTAINTY.

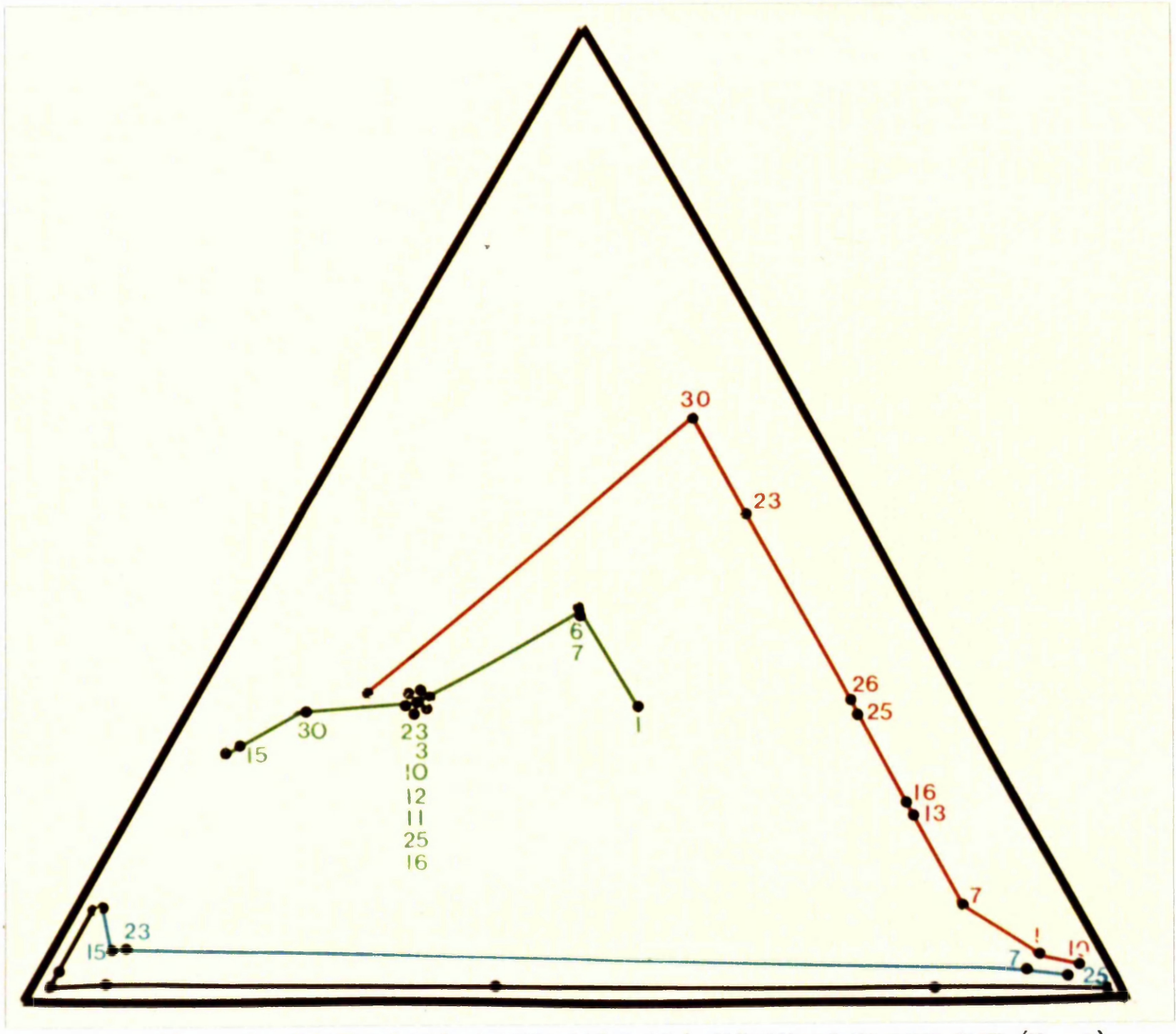


FIG. 5.9. EXAMPLE OF DIAGNOSTIC PATHS OF 3 CLINICIANS IN SAME CASE (No. 9).

reaches to the level of the triangle only at the vertices. Fig. 5.8 shows the uncertainty countours (Schmitt, 1969) in the same way as a contour map shows ocean depth. As he collects information the clinician moves around the bowl attempting to climb towards one of the vertices and so to arrive at a firm diagnosis.

Illustration of the equilateral triangle.

Fig. 5.9 shows the widely different diagnostic paths taken by three clinicians (Nos. 3, 5, 6) when dealing with the same case of thyroid carcinoma; also shown by a broken line is the path taken by the computer program. For each path the point labelled by 0 is the starting position. For a clinician this starting point corresponds to his view of the incidence rate of the three diseases in the clinic; for the computer program this starting point corresponds to the observed incidence paths in the thyroid clinic, namely 0.10, 0.89, .01 for Hashimoto's disease, simple goitre and thyroid carcinoma. The successive points marked along a route show the positions moved to after successive tests. The serial numbers of the tests selected were as follows:

Decision-maker	Tests in order selected
Clinician 3	30, 23, 26, 25, 16, 13, 7, 1, 10
Clinician 5	15, 23, 7, 25
Clinician 6	15, 30, 23, 3, 10, 12, 11, 25, 16, 6, 7, 1
Computer program	1, 9, 28, 25, 30, 7

The cluster of points on the diagnostic path of Clinician 6 corresponds to his position after tests 23, 3, 10, 12, 11, 25 and 16; in other words, having arrived there after test 23 he made no alteration to his probability assessment for the next six tests he selected. In each path the final point shown is that at which the diagnosis of thyroid carcinoma was correctly made in each case.

5.6 Discrepancies between computer and clinicians

The plotting of such paths clearly indicates a picture of broad variation between clinicians, and between some clinicians and the computer program. We can obtain, however, a more penetrating analysis by forcing a comparison between clinician and computer in each diagnostic cycle. To allow a fair comparison the computer program is made to operate with the clinician's probabilities at the start of each cycle. In each such cycle there are then two ways in which differences between the clinician's behaviour and the computer program/

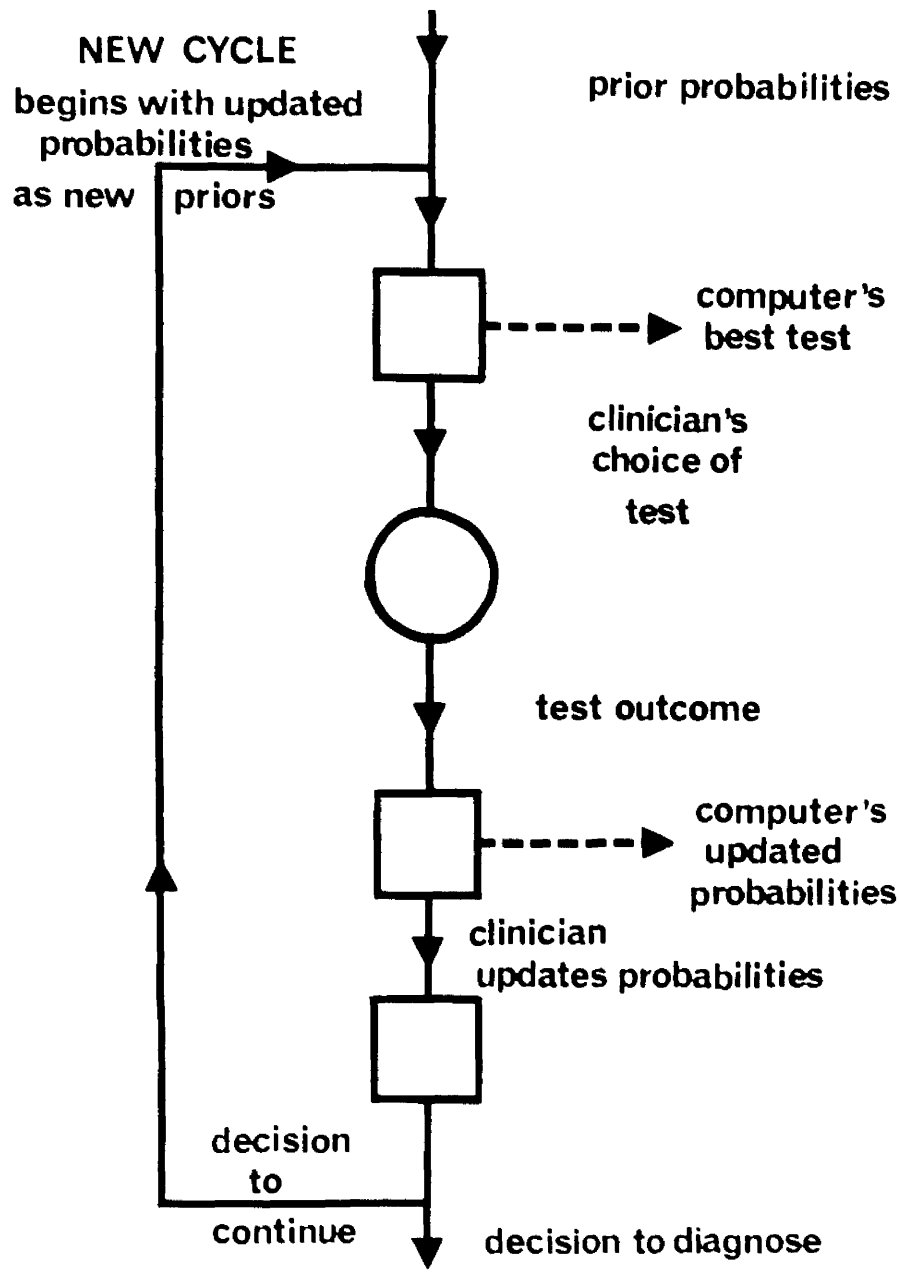


FIG. 5.10. DECISION CYCLE AND POINTS OF DISCREPANCY.

program may arise:

(1) the test chosen by the clinician may differ from the computer selection of test

(2) the updating of the probabilities as stated by the clinician may differ from the updating made by the computer. Fig. 5.10 shows a diagrammatic representation of the cycle and possible points of discrepancy.

At each cycle, and for each clinician, four measures may be computed to assess various aspects of these differences.

5.6.1. Measures of discrepancy

Test-selection discrepancy, T. At a particular stage of the diagnostic process the value of any test can be measured in terms of the amount of information it then provides; that is, the amount of uncertainty it removes. At any stage of diagnosis it is therefore natural to select the test which promises, or is expected, to provide most additional information to discriminate between the three diseases. The computer program does this (Assumption 3, Fig. 5.1). The amount by which this maximum expected gain of information exceeds the corresponding gain associated with the test chosen by the clinician is termed the test-selection discrepancy T.

Inference discrepancy, I. A measure of the discrepancy between the clinician's interpretation of a test outcome and the computer interpretation must be some overall measure of the differences between the two sets of updated probabilities. One appropriate measure is the Küllback-Leibler (1951) measure of discrimination, which we denote here by I. This has the property that it is zero when the sets of updated probabilities coincide, that is, when there is agreement between the clinician and the computer program, and is positive when there is any disagreement. The greater the disagreement the greater is the value of I.

Conservativism-liberalism index, L. In Fig. 5.4 starting from the attained position a, suppose that the computer move goes to b, while the clinician's move takes him to c. The movement from a to c represents a smaller reduction in uncertainty than the move from a to b. The clinician may therefore be said to be acting conservatively in his use of data relative to the computer program in this case. If, on the other hand, he moves to c, a position of less uncertainty than b, he is over-using the data or acting liberally relative to the computer program. For other configurations of a, b, c similar arguments apply. A/

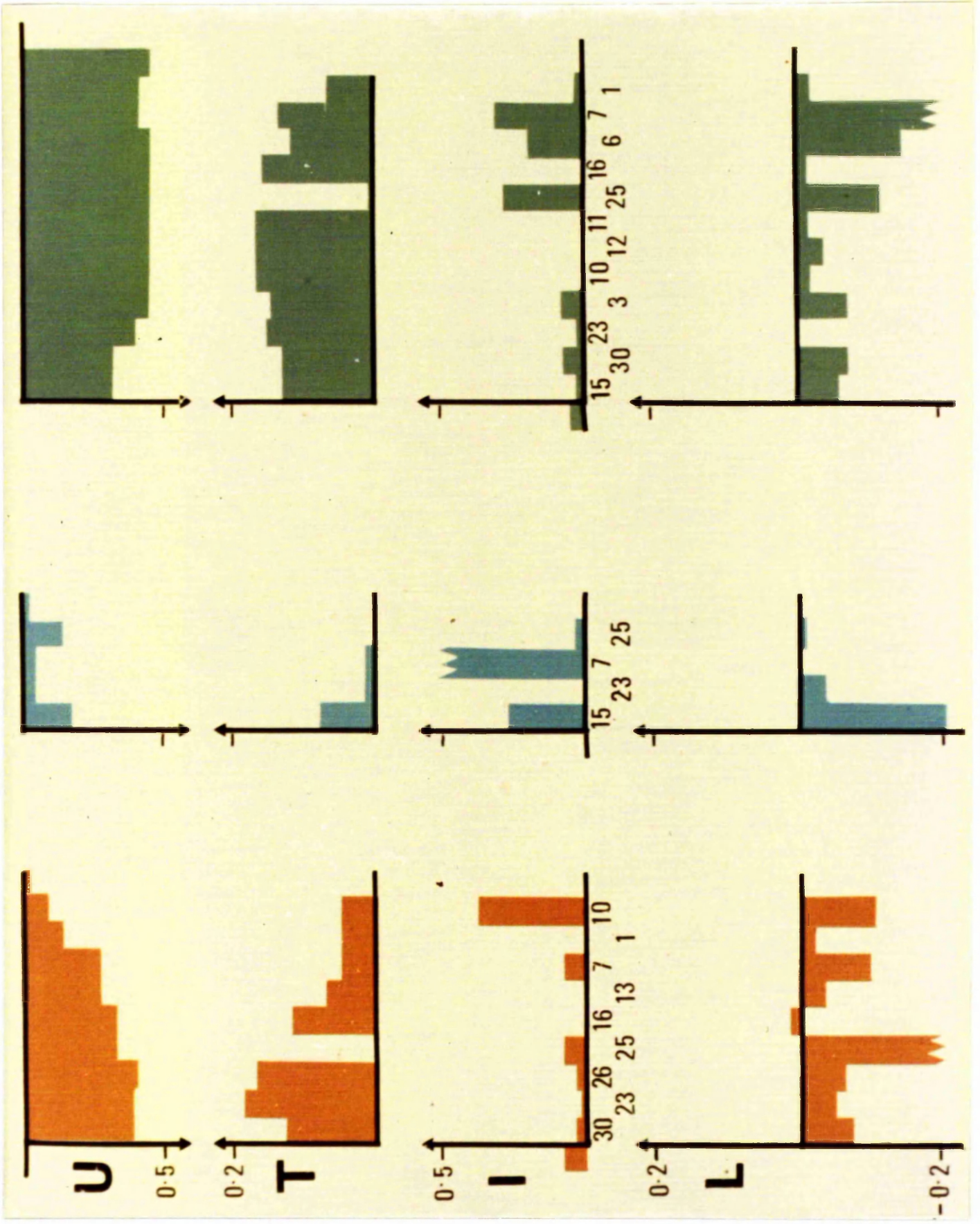


FIG. 5.11. EXAMPLE OF CYCLE-BY-CYCLE PROFILE FOR CASE EXAMPLE SHOWN IN FIG. 5.9.

A measure of this type of discrepancy is the difference L in the degrees of uncertainty associated with b and c, taken positively for liberal, and negatively for conservative behaviour.

Headway towards correct diagnosis, D. This is a simple measure to show, separately for each cycle, the relative progress of clinician and computer program towards the correct diagnosis. It is simply calculated as the difference between the updated probabilities assigned to the known correct disease, taken positively if the clinician's assessment is the larger.

5.7 Illustrative examples

Fig. 5.10 has already shown the absolute differences between clinicians and the computer program. Fig. 5.11 shows the corresponding cycle-by-cycle discrepancies and provides the (U.T.I.L.D) profile for each clinician in the same case of thyroid carcinoma. The effectiveness of this form of analysis can be illustrated by some observations on these results. It must be emphasised that no value judgements are implied in these comparisons. The computer program merely provides an objective indirect comparison among a number of clinicians.

(a)/

(a) Clinician 3 shows a large discrepancy in test selection (T) but is close to the computer program in interpreting these tests (I). He is generally conservative in his use of data ($L < 0$), but is mildly liberal in the use of information from test 16.

(b) Clinician 5 reduces the degree of uncertainty (U) more rapidly than the others. He is also much nearer the computer program in test selection (T) and uses fewer tests than clinicians 3 and 6 to reach a correct diagnosis. In test 7 he is liberal ($L > 0$) in probability estimation and is ahead of the normative model in progress towards the diagnosis (D).

(c) Clinician 6 is more uncertain at the end of his case than when he began (U); he is consistently conservative ($L < 0$) and shows a large discrepancy in test selection (T). His inference discrepancy (I) is low, but this may be explained by the fact that the tests he selects provide little information when this is calculated in accordance with the computer program.

The three clinicians select test 25 at different stages of their diagnostic paths, and yet the test selection discrepancy T is zero for each. This is simply explained by test 25 being highly informative for most starting positions in the triangle.

5.8 Results

The diagnostic performance of the six clinicians in the study can be analysed at several levels.

The simplest level is that of diagnostic accuracy and of the number of tests used for each case by the clinicians.

More complex information can be acquired by the analysis of the 'route' or path taken by the clinicians to a diagnosis. Wide differences are revealed by this analysis, particularly if the 'entropy' measure (Shannon, 1948) is used.

Differences within the group are more easily demonstrated with the four measures of discrepancy which, with the 'entropy' measure, make up the diagnostic profile.

Interpretation of individual difference is greatly helped by the addition of data from personality tests and simple information about clinical experience.

5.8.1 Diagnostic accuracy

Cases of varying difficulty were selected for use in the study by the objective technique already described (Fig. 4.5). Two of these cases were misdiagnosed by the computer model, giving it an overall accuracy of 18/20 in the study.

The accuracy of the six clinicians varied from 17 to

20/

	Clinicians ranked for obsessionalism					
	1	6	5	4	2	3
Number of correct diagnoses out of 20*	20	20	19	18	18	17
Total number of tests used	213	174	141	141	120	153

* Computer 18/20

FIG. 5.12. DIAGNOSTIC ACCURACY IN STUDY.

			CORRECT DIAGNOSIS					
Case	Computer	Diagnosis	1	2	3	4	5	6
1	2	S						
2A	3	S						
2B	3	S						
4	3	S						
9	16	C						
11A	.98	S						
11B	.98	S						
12	10	C					H	
13	5	H		S	S	S		
15	2	S						
16	2	S						
17A	.6	S						
17B	16	S						
20	2	S						
23A	2	S						
23B	2	S						
24	XC	H		C	C	S		
25	2	S						
28	XC	S			C			
36	3	S						
			20	18	17	18	19	20

FIG. 5.13. DIAGNOSTIC ACCURACY WITH WRONG DIAGNOSES IN BOXES: SECOND COLUMN GIVES (a) NUMBER OF TESTS IF CORRECT (b) FINAL PROBABILITY IF <.99 and (c) WRONG DIAGNOSIS OF CARCINOMA in 24 and 28, FOR COMPUTER.

20 (Fig. 5.12). An analysis of individual cases (Fig. 5.13) shows that in 16/20 cases all clinicians were correct; in no case were more than 3 clinicians wrong.

It is important to note that in 20 per cent of cases the final probability in the chosen diagnosis was <.8. This is reflected in the measure of "average uncertainty before diagnosis" and seems to be related to obsessionalism (5.8.6).

In the two cases which were misdiagnosed by the computer, 3 out of 6 (No. 24) and 5 out of 6 clinicians (No. 28) were correct in their diagnosis (Fig. 5.13). This means that the computer, making use of Bayes' theorem and all the available data (30 tests) was less effective than the clinicians using between 6 and 11 tests per case. Since it is unlikely that all the clinicians were guessing, it suggests that the successful clinicians were processing the data in a more effective way than the computer. They may well be using some pattern of interdependence among the data which the computer program, with its assumption of independence (2.1.6) has ignored.

It is clearly of interest to be able to detect and study in detail such 'pattern recognition'. Later in this chapter a method (based on the 'liberalism' and 'diagnosis' components/

NUMBER OF TESTS USED

Case	1	2	3	4	5	6	Computer
1	10	3	5	6	6	6	2
2A	7	6	4	4	5	12	3
2B	11	4	8	8	7	12	3
4	12	8	9	4	8	9	3
9	9	6	9	6	4	12	16
11A	10	7	8	7	6	10	.98
11B	11	10	8	7	6	13	.98
12	11	7	8	6	8	9	10
13	11	15	12	9	10	11	5
15	15	5	8	10	7	6	2
16	11	4	6	7	7	7	2
17A	10	4	6	7	7	7	.6
17B	11	4	7	7	7	5	.6
20	11	3	8	7	7	7	2
23A	12	4	6	7	7	8	2
23B	10	4	5	6	7	6	2
24	9	10	8	10	11	11	XC
25	11	4	4	8	7	6	2
28	10	7	12	6	7	11	XC
36	11	5	12	9	7	6	3
A	214	120	153	141	141	174	237
B	153	78	104	97	97	117	57

FIG.5.14. NUMBER OF TESTS FOR A (ALL CASES) AND B (WITHOUT THOSE IN BOXES).

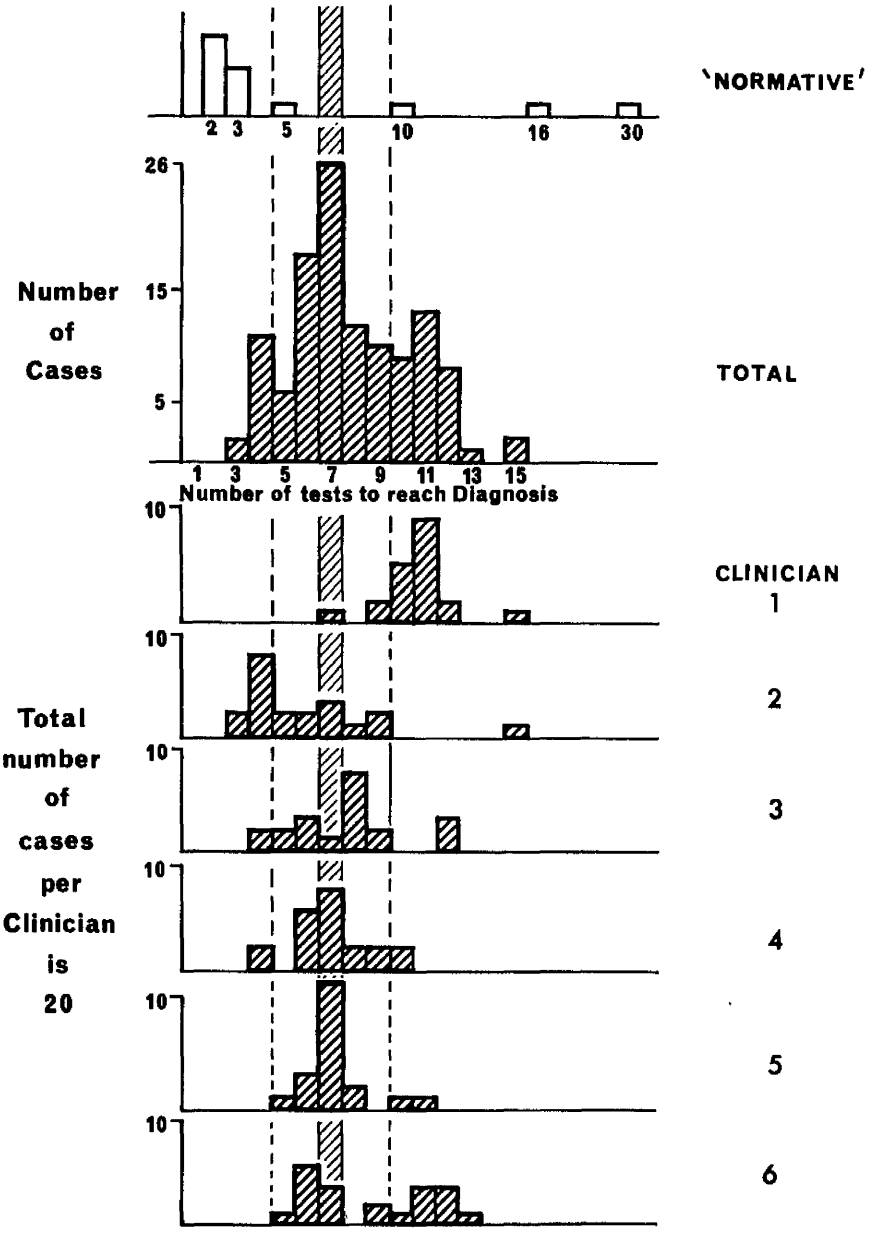


FIG. 5.15. TOTAL NUMBER OF TESTS AS HISTOGRAMS.

components of the 'profile') is suggested which may help to detect the more obvious instances of 'pattern recognition' in studies such as this.

5.8.2 Number of tests used

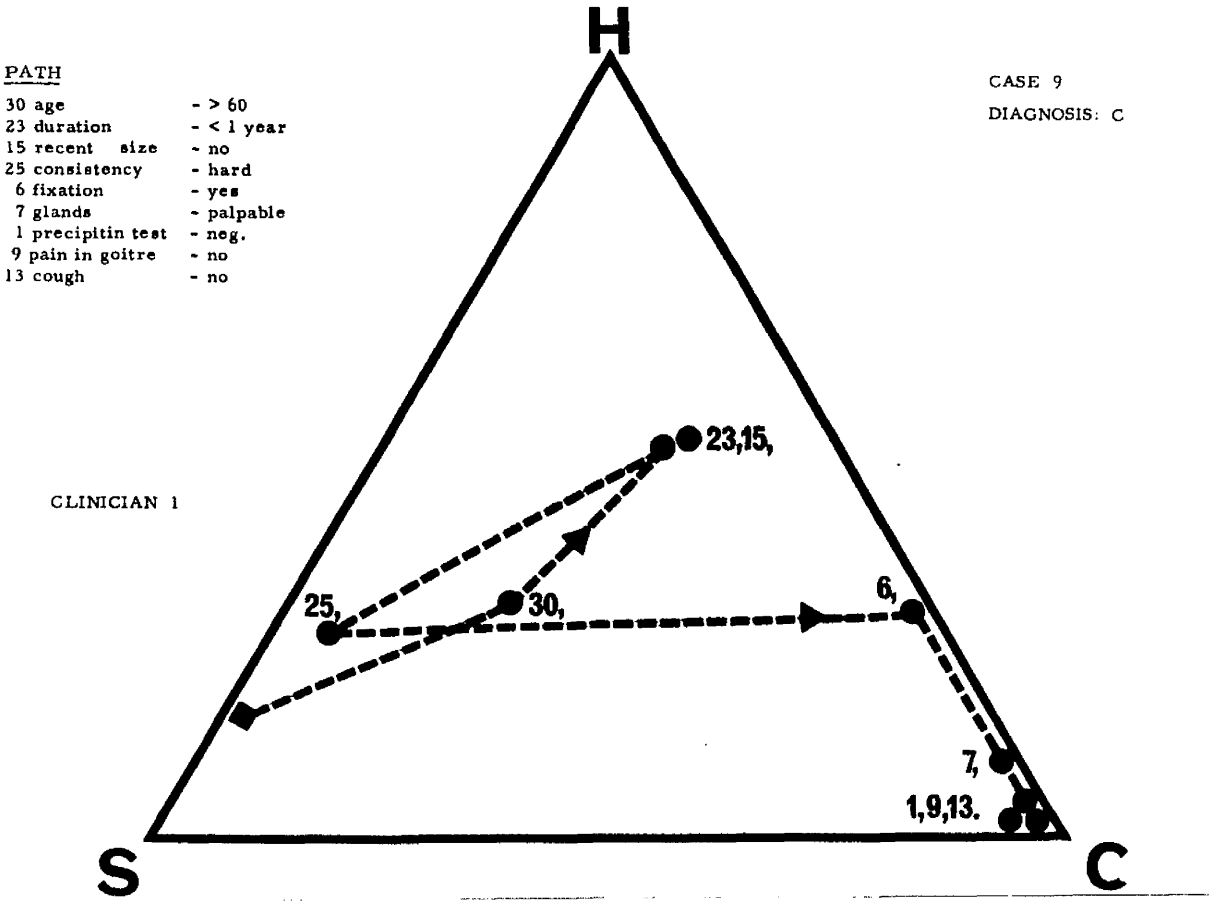
For all 20 cases in the study the computer model took 237 cases, while the clinicians varied widely from 120 to 214 (Fig. 5.14). If the 6 cases in which the computer was either wrong or did not reach .99 are removed (since all 30 tests are used in each case), the computer clearly uses many fewer tests than the clinicians. In most individual cases the computer takes fewer tests.

It is worth noting that clinician 1 took almost twice as many tests as clinician 2. This appears to be related to the degree of obsessionalism of each and is discussed in detail later in the chapter (5.8.5; 5.9.3; 5.9.6).

Another noteworth feature is that most of the cases (72/120) including the most difficult ones, were diagnosed with between 5 and 9 tests per case. This feature (summarised in Fig. 5.15) may be explained by limitations of short-term memory or by the clinician's capacity for processing information. This feature is dealt with in more detail later in this chapter (5.8.8).

case	clinician 1	2	3	4	5	6	
2	4	2	4	4	2	0	
11	1	3	0	0	0	3	
17	1	0	1	0	0	2	
23	2	0	1	1	0	2	
	8	5	6	5	2	7	total

FIG. 5.16. CONSISTENCY IN NUMBER OF TESTS USED IN THOSE CASES WHICH WERE REPEATED IN THE STUDY.



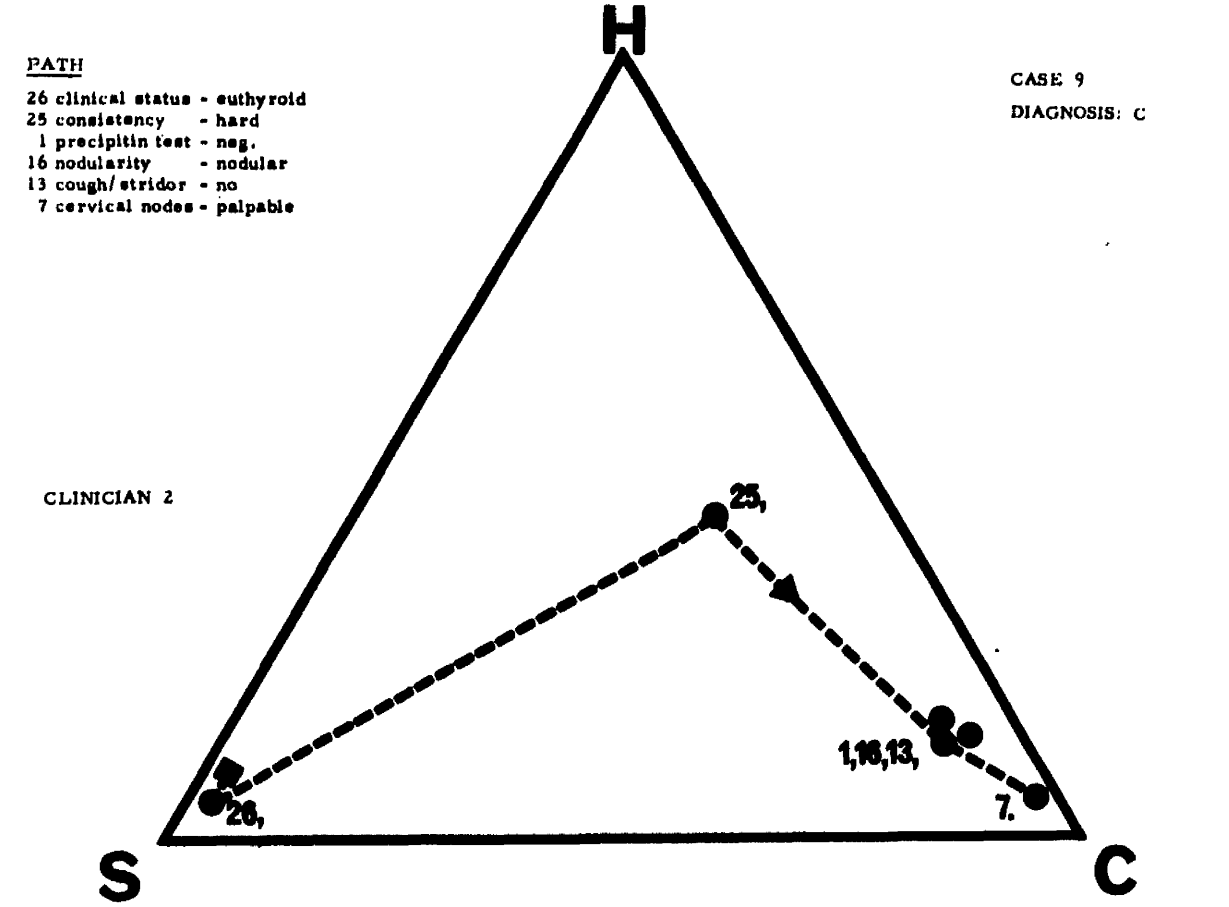


FIG. 5.17. (a) PATH OF CLINICIANS 1 AND 2 IN SAME CASE AS FIG. 5.9.

Four cases (2, 11, 17, 23) were repeated in the second half of the study. When the number of tests used in these cases is compared for each half of the study, variation is again seen (Fig. 5.16). For example, in case 2 there is a 100 per cent difference between case 2A and 2B for clinicians 3 and 4. In general, consistency (or reproducibility) as measured in this way is good and clinician 5 is the most consistent.

5.8.3 Individual paths to a diagnosis

The wide variation in the routes taken by clinicians to a diagnosis has been clearly shown in Fig. 5.9, while Fig. 5.17 shows the remaining 3 clinicians in the same case. This variability takes many forms.

Thus, clinician 6 takes seven test outcomes before altering his probability assessment and is even more uncertain after doing so. Clinician 2 hesitates and takes 3 test outcomes before making a final diagnosis. The indecisiveness of clinician 6 in this example is typical of his overall performance, while with clinician 2 the three tests are aimed at finally excluding Hashimoto's disease and confirming carcinoma.

While it is informative to compare the 6 'routes' for case 9 by inspection, it would be better if more formal methods of comparison were possible.

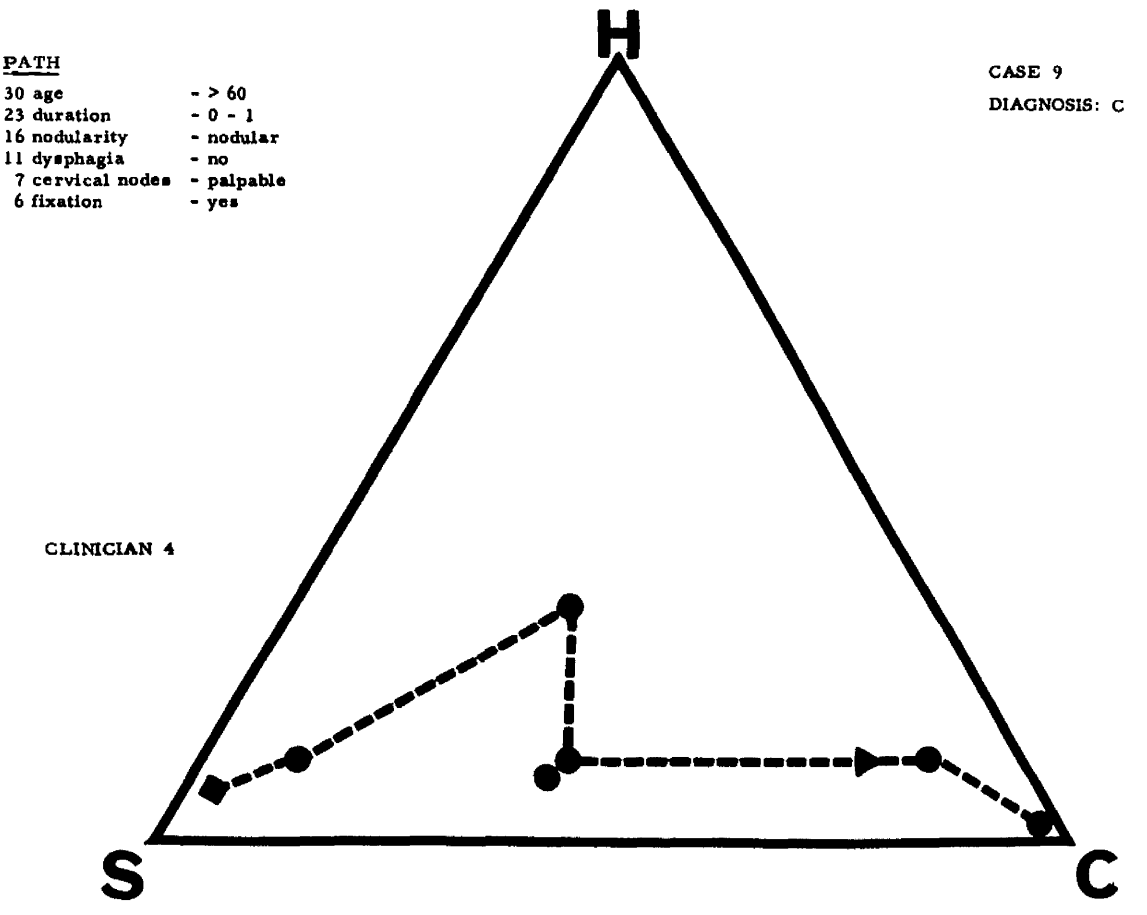


FIG. 5.17 (b) PATH OF CLINICIAN 4 IN SAME CASE AS FIG. 5.9.

CLINICIAN	AVERAGE UNCERTAINTY PER TEST	AVERAGE UNCERTAINTY BEFORE DIAGNOSIS
1	.0156	.2433
2	.0117	.1279
3	.0178	.1213
4	.0117	.0641
5	.0074	.0289
6	.0156	.2251

FIG. 5.18. VALUES FOR UNCERTAINTY MEASURES.

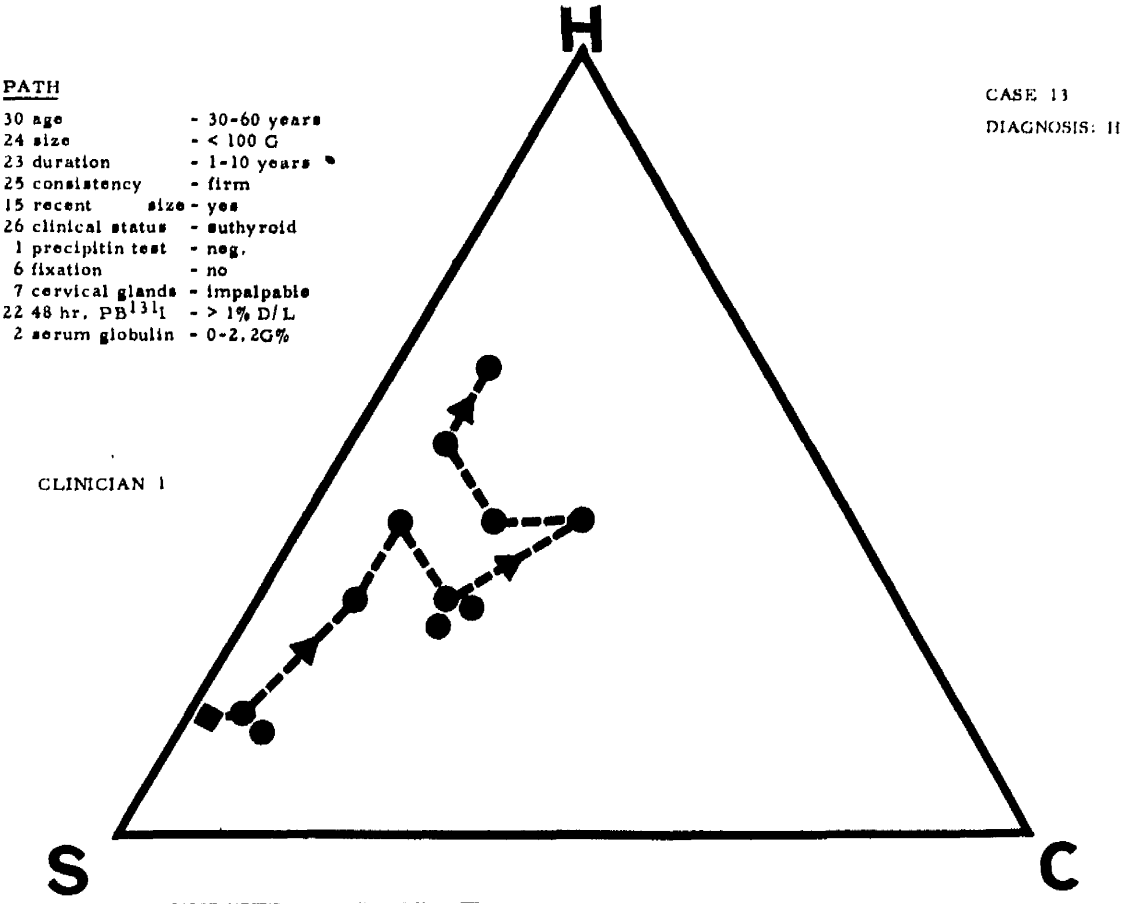
The most obvious way would be to make use of the 'uncertainty' measure (5.5) which reduces the three probabilities to a single value. It can be used in two ways; by calculating (i) the average undertainty per test, and (ii) the average uncertainty before diagnosis.

The average uncertainty per test is a guide to how long the clinician spends in the central (lower) part of the triangular bowl. It will be low if the route towards a corner is rapid and direct since there will be fewer tests in the centre where the uncertainty value is high (Fig. 5.8).

The average uncertainty before diagnosis will measure how near the clinician goes to a corner (or vertex Fig.5.7) before settling on his final diagnosis.

The values for these two measures are shown in Fig. 5.18. Clinician 5 has the lowest value for both, while clinicians 1 and 6 have high values for both. The significance of these results is best assessed in relation to personality factors and to clinical experience (see below, 5.8.6; 5.8.7).

The varied routes of the 6 clinicians in a case of Hashimoto's disease are shown in Fig. 5.19. The computer model took 5 tests and was correct, while clinicians 2, 3 and/



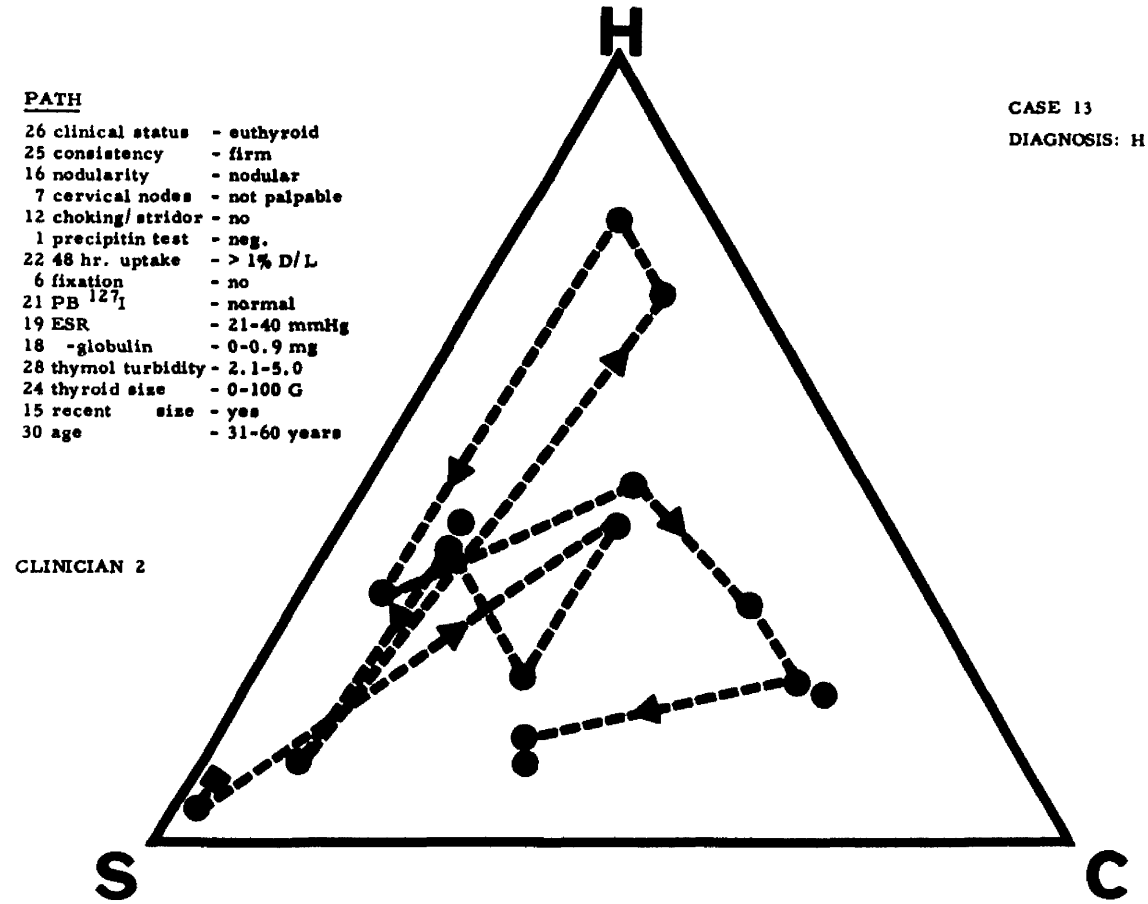
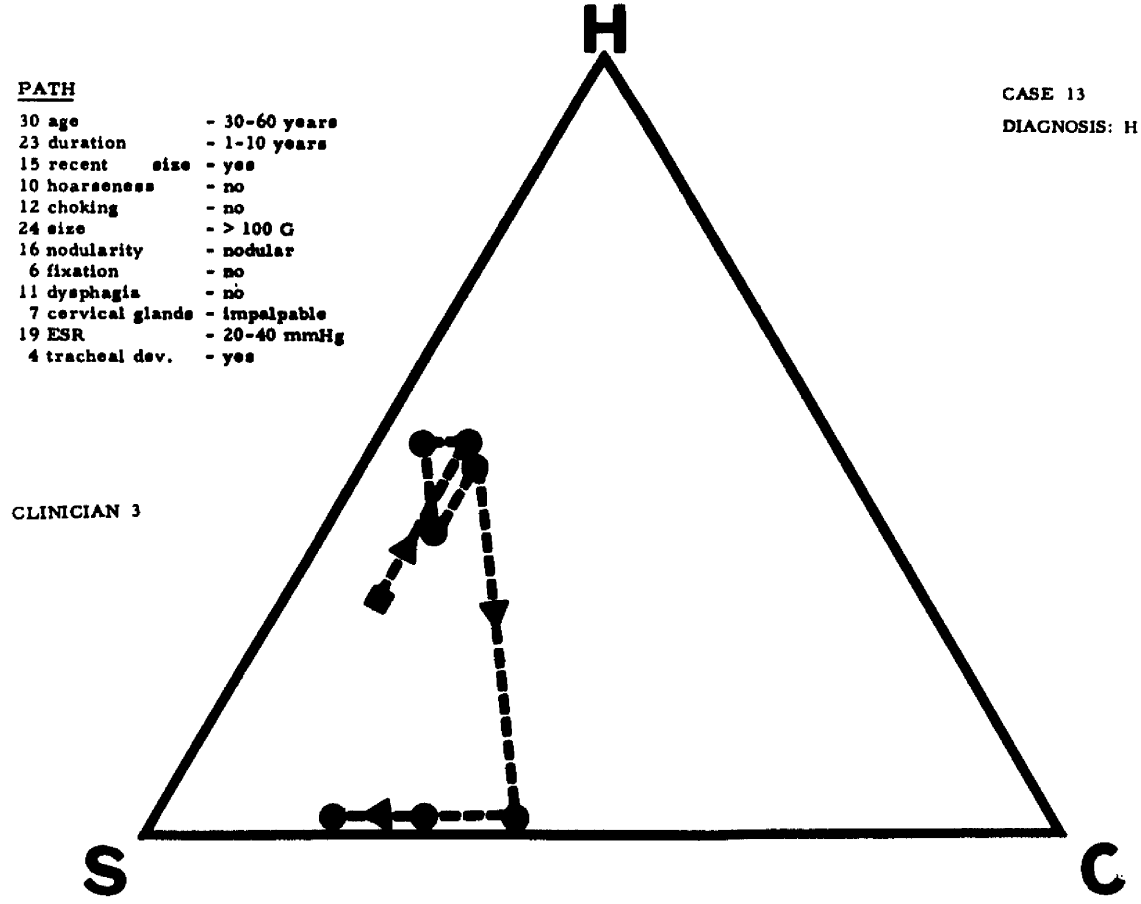


FIG. 5.19. (a) PATH OF CLINICIANS 1 AND 2 IN CASE OF HASHIMOTO'S DISEASE.



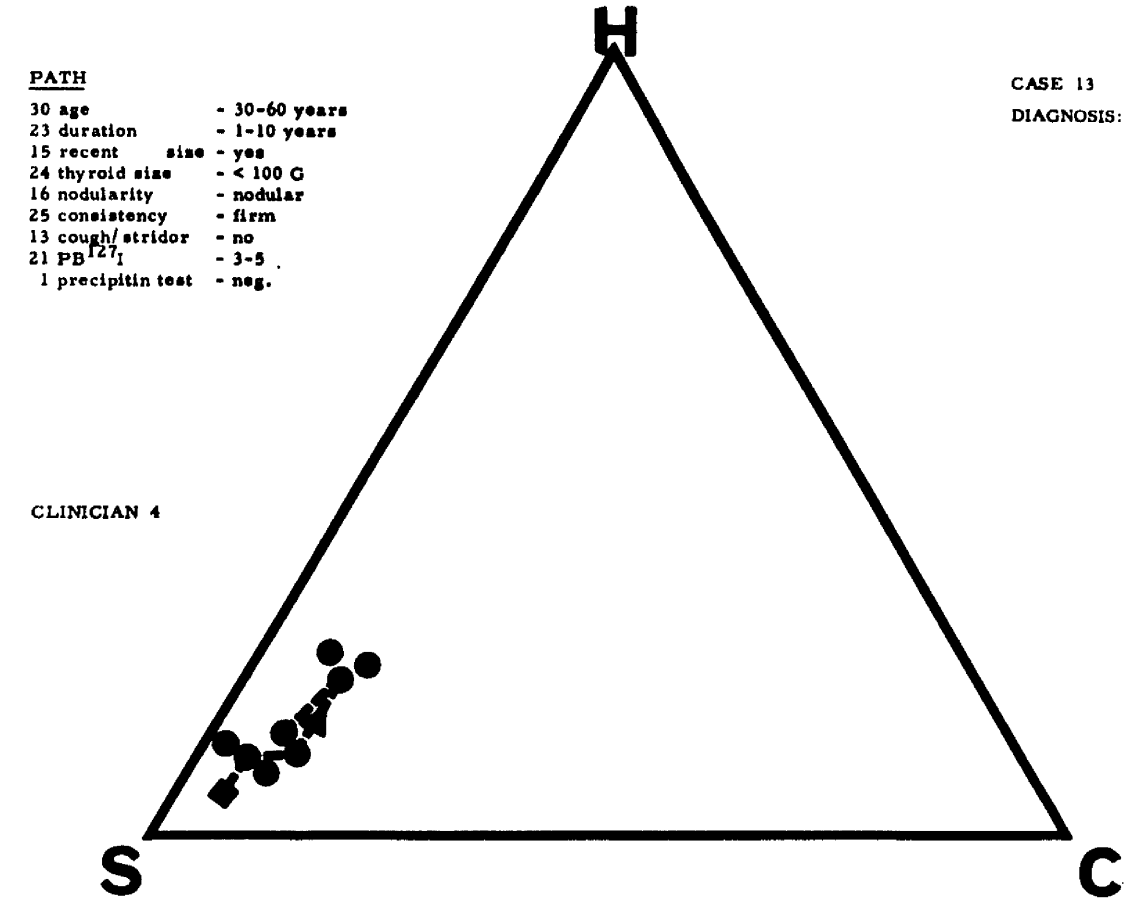


FIG. 5.19. (b) PATH OF CLINICIANS 3 AND 4 IN CASE OF HASHIMOTO'S DISEASE.

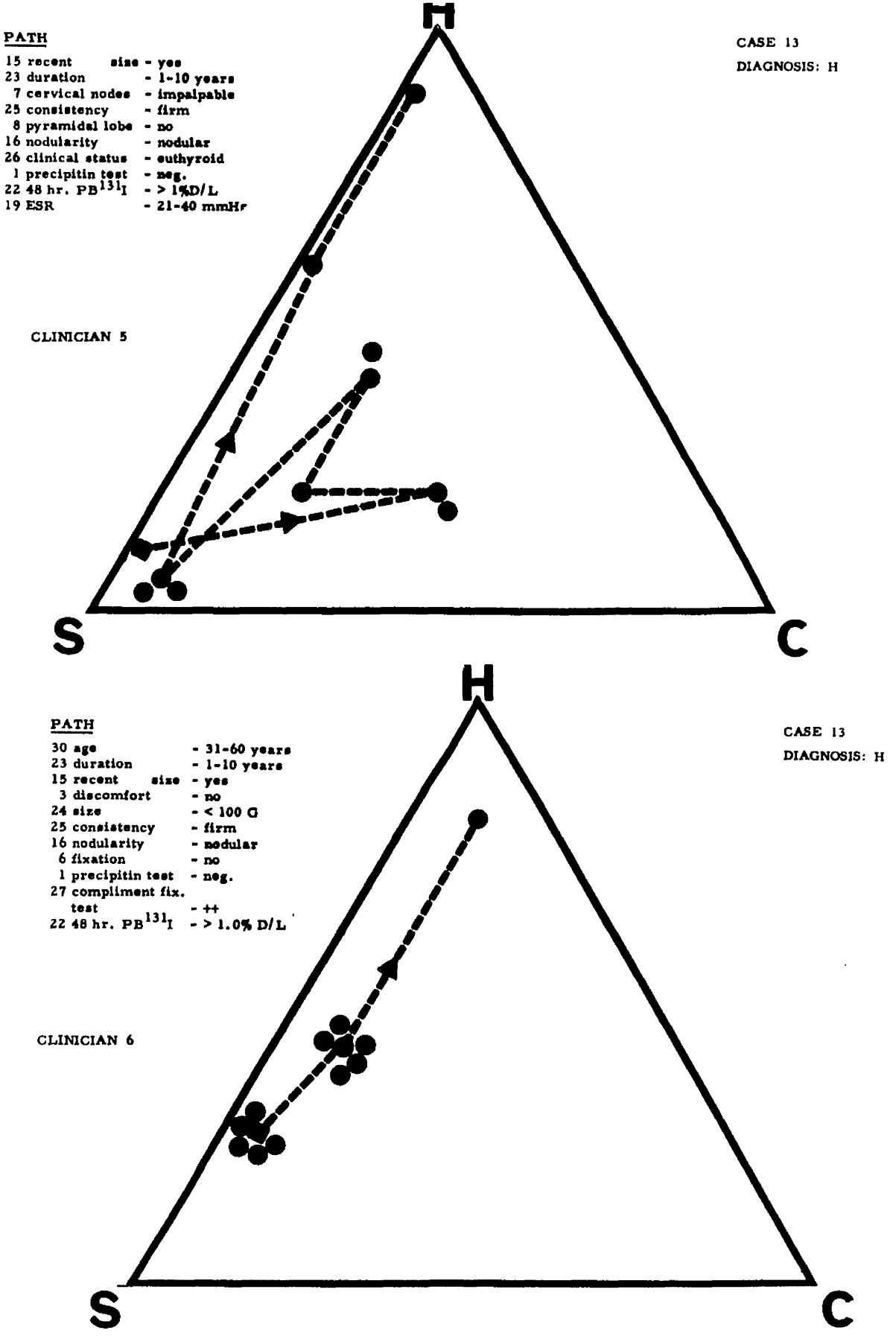


FIG. 5.19. (c) PATH OF CLINICIANS 5 and 6 IN CASE OF HASHIMOTO'S DISEASE.

and 4 misdiagnosed the case as one of simple goitre (Fig.5.13). Clinician 2 in particular has a very variable route before making a wrong diagnosis.

In the same case clinician 6, although slow to make use of the test outcomes which he has asked for, moves deliberately to the correct diagnosis. Clinician 5 after noting that the gland is firm (test 25) is deflected from Hashimoto's disease by the absence of a pyramidal lobe. He then moves to eliminate the possibility of simple goitre (tests 16, 26, 1) before opting decisively for Hashimoto's disease.

The most striking features of all in these illustrative examples is the individuality and variability in the diagnostic styles of the clinicians when tackling identical problems.

5.8.5 'Profile analysis': measures of discrepancy

The first element of the 'profile' is the uncertainty measure and this has been dealt with in some detail in the preceding section (5.8.4). For the remaining measures (test selection, inference, conservativism/liberalism and diagnosis), the average values per test are shown in Fig. 5.20. When these values are ranked and compared with the measures of obsessionalism, extroversion and clinical experience/

CLINICIAN 1

TEST	NO.	T	I	LIB.	CON.	D
1	11	.0361	.0259	.0018	-.0642	-.0291
2	2	.1315	.0334	.0000	-.1041	-.1601
3	0	.0000	.0000	.0000	.0000	.0000
4	0	.0000	.0000	.0000	.0000	.0000
5	1	.1024	.0022	.0000	-.0289	.0378
6	18	.1052	.0149	.0006	-.0325	-.0344
7	18	.1076	.0081	.0076	-.0279	-.0222
8	0	.0000	.0000	.0000	.0000	.0000
9	8	.0959	.0024	.0000	-.0130	-.0023
10	2	.1287	.0065	.0000	-.0301	.0017
11	3	.1096	.0007	.0000	-.0200	-.0203
12	1	.1629	.0084	.0007	.0000	.0621
13	3	.1235	.0021	.0000	-.0142	-.0117
14	1	.1744	.0054	.0000	-.0063	.0761
15	17	.1065	.0441	.0085	-.0392	-.0264
16	9	.1188	.0055	.0027	-.0362	-.0206
17	0	.0000	.0000	.0000	.0000	.0000
18	0	.0000	.0000	.0000	.0000	.0000
19	1	.0567	.0168	.0000	-.0919	-.1123
20	11	.1232	.0207	.0294	-.0214	.0450
21	2	.0483	.0114	.0000	-.0241	.0806
22	17	.0653	.0283	.0001	-.0900	-.1119
23	20	.1293	.0139	.0123	-.0220	-.0213
24	16	.1068	.0870	.0036	-.0755	.1509
25	20	.0410	.0762	.0000	-.1570	-.1138
26	13	.1127	.0037	.0059	-.0103	-.0050
27	0	.0000	.0000	.0000	.0000	.0000
28	0	.0000	.0000	.0000	.0000	.0000
29	0	.0000	.0000	.0000	.0000	.0000
30	20	.1056	.0396	.0575	-.0677	-.1032

FIG. 5.20. (a) AVERAGE VALUES FOR PROFILE MEASURES FOR CLINICIAN 1.

CLINICIAN 2

TEST	NO	T	I	LIB.	CON.	D
1	13	.0160	.0190	.0172	-.0153	-.0164
2	0	.0000	.0000	.0000	.0000	.0000
3	0	.0000	.0000	.0000	.0000	.0000
4	0	.0000	.0000	.0000	.0000	.0000
5	0	.0000	.0000	.0000	.0000	.0000
6	8	.0797	.0383	.0304	-.0218	.1191
7	10	.0876	.0222	.0044	-.0334	.0300
8	0	.0000	.0000	.0000	.0000	.0000
9	0	.0000	.0000	.0000	.0000	.0000
10	1	.1637	.0016	.0000	-.0163	.0255
11	0	.0000	.0000	.0000	.0000	.0000
12	2	.1814	.0162	.0000	-.0695	.0800
13	2	.1236	.0113	.0376	-.0205	.0924
14	0	.0000	.0000	.0000	.0000	.0000
15	3	.1290	.0614	.0520	-.0170	-.0107
16	7	.1315	.0159	.0049	-.0189	-.0063
17	1	.1276	.0197	.0374	.0000	.0921
18	1	.1144	.0068	.0000	-.0175	.0225
19	1	.0785	.1869	.0000	-.2844	-.3776
20	11	.0932	.0369	.0203	-.0025	-.0334
21	5	.0186	.0455	.0079	-.0136	-.0737
22	11	.0546	.0254	.0053	-.0227	-.0572
23	0	.0000	.0000	.0000	.0000	.0000
24	2	.1471	.0585	.0000	-.0639	.1797
25	20	.0146	.1479	.0193	-.1189	-.0942
26	20	.0593	.0024	.0333	.0000	.0176
27	0	.0000	.0000	.0000	.0000	.0000
28	1	.0662	.1486	.0000	-.0529	-.3573
29	0	.0000	.0000	.0000	.0000	.0000
30	1	.0821	.0122	.0000	-.0123	-.0557

FIG. 5.20. (b) AVERAGE VALUES FOR PROFILE MEASURES FOR CLINICIAN 2.

CLINICIAN 3

TEST NO.	T	I	LIB.	CON.	D
1 2	.0395	.1517	.0152	-.0116	-.0230
2 0	.0000	.0000	.0000	.0000	.0000
3 1	.1541	.0083	.0000	-.0035	-.0158
4 3	.0368	.2578	.0027	-.0144	.2007
5 0	.0000	.0000	.0000	.0000	.0000
6 9	.0820	.1093	.0432	-.0181	.0258
7 5	.0701	.0465	.0197	-.0275	-.0114
8 2	.1605	.3994	.1698	-.0050	.1296
9 0	.0000	.0000	.0000	.0000	.0000
10 5	.1365	.0903	.0000	-.0499	-.0127
11 3	.1740	.0026	.0094	-.0121	-.0042
12 4	.1737	.0449	.0198	-.0102	-.0144
13 3	.1527	.0030	.0109	-.0137	.0463
14 0	.0000	.0000	.0000	.0000	.0000
15 11	.1362	.0896	.0131	-.0499	-.0622
16 18	.1569	.3006	.0172	-.0352	.0160
17 0	.0000	.0000	.0000	.0000	.0000
18 0	.0000	.0000	.0000	.0000	.0000
19 3	.0574	.4827	.0324	-.1462	.1528
20 1	.0695	1.3808	.0000	-.2346	-.2308
21 0	.0000	.0000	.0000	.0000	.0000
22 0	.0000	.0000	.0000	.0000	.0000
23 18	.1604	.0242	.0075	-.0263	-.0215
24 6	.1540	.3652	.0002	-.1284	.0200
25 19	.0395	.0779	.0015	-.1647	-.1564
26 19	.1335	.0460	.0220	-.0193	.0555
27 0	.0000	.0000	.0000	.0000	.0000
28 1	.0335	.0185	.0707	.0000	-.0549
29 0	.0000	.0000	.0000	.0000	.0000
30 20	.1328	.0417	.0005	-.0779	-.0946

FIG. 5.20. (c) AVERAGE VALUES FOR PROFILE MEASURES FOR CLINICIAN 3.

CLINICIAN 4

TEST	NO.	T	I	LIB.	CON.	D
1	3	.0234	.1042	.0000	-.0573	-.0103
2	0	.0000	.0000	.0000	.0000	.0000
3	0	.0000	.0000	.0000	.0000	.0000
4	0	.0000	.0000	.0000	.0000	.0000
5	0	.0000	.0000	.0000	.0000	.0000
6	4	.1410	.1718	.0639	-.0118	.0291
7	5	.1044	.0599	.0411	-.0229	-.0289
8	0	.0000	.0000	.0000	.0000	.0000
9	0	.0000	.0000	.0000	.0000	.0000
10	0	.0000	.0000	.0000	.0000	.0000
11	7	.0893	.0722	.0161	-.0250	.0263
12	5	.0831	.3003	.0045	-.0684	.0829
13	2	.1337	.0034	.0000	-.0300	.0224
14	0	.0000	.0000	.0000	.0000	.0000
15	17	.0808	.1079	.0247	-.0248	-.0291
16	11	.0976	.1011	.0105	-.0856	.0010
17	0	.0000	.0000	.0000	.0000	.0000
18	0	.0000	.0000	.0000	.0000	.0000
19	0	.0000	.0000	.0000	.0000	.0000
20	0	.0000	.0000	.0000	.0000	.0000
21	1	.0534	.0079	.0000	-.0072	.0510
22	0	.0000	.0000	.0000	.0000	.0000
23	20	.0843	.0102	.0126	-.0304	.0123
24	17	.0824	.3547	.0211	-.1068	.1820
25	14	.0139	.0543	.0019	-.1424	-.0679
26	12	.0709	.2834	.0835	-.0083	.0411
27	0	.0000	.0000	.0000	.0000	.0000
28	0	.0000	.0000	.0000	.0000	.0000
29	0	.0000	.0000	.0000	.0000	.0000
30	20	.0431	.0442	.0000	-.1270	-.1117

FIG. 5.20. (d) AVERAGE VALUES FOR PROFILE MEASURES FOR CLINICIAN 4.

CLINICIAN 5

TEST	NO.	T	I	LIB.	CON.	D
1	12	.0082	.0185	.0246	-.0244	.0243
2	1	.1034	.0200	.0000	-.1146	-.1047
3	0	.0000	.0000	.0000	.0000	.0000
4	0	.0000	.0000	.0000	.0000	.0000
5	1	.0423	.0015	.0000	-.0308	.0258
6	1	.0757	.0098	.0000	-.0621	-.0580
7	19	.0440	.0869	.0146	-.0117	.0725
8	3	.1376	.0519	.0694	-.0010	.1328
9	0	.0000	.0000	.0000	.0000	.0000
10	0	.0000	.0000	.0000	.0000	.0000
11	0	.0000	.0000	.0000	.0000	.0000
12	0	.0000	.0000	.0000	.0000	.0000
13	0	.0000	.0000	.0000	.0000	.0000
14	0	.0000	.0000	.0000	.0000	.0000
15	19	.0752	.0592	.0558	-.0408	.0279
16	1	.1842	.3631	.0000	-.2773	-.3322
17	1	.0599	.0517	.1116	.0000	.0711
18	0	.0000	.0000	.0000	.0000	.0000
19	7	.0426	.2969	.0456	-.0627	.0893
20	0	.0000	.0000	.0000	.0000	.0000
21	0	.0000	.0000	.0000	.0000	.0000
22	18	.0132	.0406	.0091	-.0252	.0188
23	20	.0828	.0060	.0006	-.0324	-.0010
24	0	.0000	.0000	.0000	.0000	.0000
25	20	.0243	.0420	.0052	-.0700	.0117
26	16	.0282	.0013	.0000	-.0091	.0006
27	0	.0000	.0000	.0000	.0000	.0000
28	1	.0513	.0818	.0000	-.1653	.2736
29	0	.0000	.0000	.0000	.0000	.0000
30	1	.0981	.0420	.0000	-.0008	.0455

FIG. 5.20. (e) AVERAGE VALUES FOR PROFILE MEASURES FOR CLINICIAN 5.

CLINICIAN 6

TEST	NO.	T	I	LIB.	CON	D
1	6	.0313	.1318	.0000	-.0838	.1812
2	0	.0000	.0000	.0000	.0000	.0000
3	5	.1454	.0408	.0000	-.0345	.0196
4	1	.0499	.0533	.0000	-.1612	-.2261
5	1	.0967	.0033	.0000	-.0027	-.0401
6	10	.1035	.0450	.0081	-.0402	-.0903
7	8	.1133	.0861	.0000	-.0541	-.1601
8	0	.0000	.0000	.0000	.0000	.0000
9	1	.1609	.0010	.0000	-.0108	-.0008
10	3	.1273	.0058	.0000	-.0217	-.0110
11	2	.1322	.0009	.0000	-.0140	.0067
12	1	.1689	.0085	.0000	-.0353	.0736
13	0	.0000	.0000	.0000	.0000	.0000
14	0	.0000	.0000	.0000	.0000	.0000
15	19	.1185	.0444	.0039	-.0425	.0521
16	18	.1030	.0057	.0036	-.0308	-.0052
17	0	.0000	.0000	.0000	.0000	.0000
18	0	.0000	.0000	.0000	.0000	.0000
19	0	.0000	.0000	.0000	.0000	.0000
20	5	.0785	.0372	.0000	-.0586	.0820
21	1	.0361	.0039	.0000	-.0305	.0371
22	10	.0339	.0225	.0016	-.0388	-.0621
23	20	.1384	.0144	.0031	-.0361	-.0026
24	18	.1208	.1525	.0054	-.0814	.2005
25	20	.0375	.0515	.0001	-.1099	-.0688
26	2	.1141	.0036	.0000	-.0072	-.0249
27	3	.0000	.1237	.0000	-.1568	-.2156
28	0	.0000	.0000	.0000	.0000	.0000
29	0	.0000	.0000	.0000	.0000	.0000
30	20	.1302	.0385	.0069	-.0667	-.0471

FIG. 5.20. (f) AVERAGE VALUES FOR PROFILE MEASURES FOR CLINICIAN 6.

RANK ORDER FOR PROFILE MEANS AND CLINICAL EXPERIENCE

Obsessionalism (16 P.F.)	Experience G.	Experience T.	Test Selection	Inference	Lib.	Cons.	Average 'U' per test	Average 'U' before diagnosis
1	5	5	3	3	6	5	3	1
6	4	2	1	4	1	3	1	6
5	2	6	6	5	2	6	6	2
4	6	1	2	2	4	4	4	3
2	1	4	4	6	5	1	2	4
3	3	3	5	1	3	2	5	5

FIG. 5.21. RELATIONSHIP OF RANKED VALUES FOR PROFILE, PERSONALITY AND EXPERIENCE MEASURES.

experience a number of possible associations emerge (Fig.5.21). Thus, clinical experience appears to be related to test selection, while conservativism is correlated with neurotic obsessionalism (5.8.6). These findings will be discussed in detail later (5.9.6; 5.9.7).

The inclusion of the 'diagnosis' measure in the profile was primarily aimed at detecting unusually effective use of data. Thus if a clinician uses data to move to the correct diagnosis much faster than the computer, then this may indicate the use of something like 'pattern recognition'.

By combining the liberalism and the diagnosis measures it may be possible to differentiate 'guessing' from such pattern recognition. This type of analysis will be disucssed in more detail later under the heading of 'pattern recognition' (5.9.9).

Another feature which was investigated was the relationship between the test selection and conservativism/ liberalism measures and the frequency with which tests were selected.

When test selection values are considered there is a clear tendency to select effectively (i.e. 'T' is low) those tests which are chosen frequently. Conversely, tests which are seldom selected tend to have a high 'T' value (Fig. 5.22).

"TEST SELECTION"

	TOP TWO				BOTTOM TWO			
	T	F	T	F	T	F	T	F
1	1	11	25	20	14	1	12	1
2	25	20	1	13	12	2	10	1
CLINICIAN 3	28	1	29	19	11	3	12	4
4	25	14	1	3	6	4	13	2
5	1	12	22	18	16	1	8	3
6	1	6	22	10	12	1	9	1
	TOTAL	147			TOTAL	24		

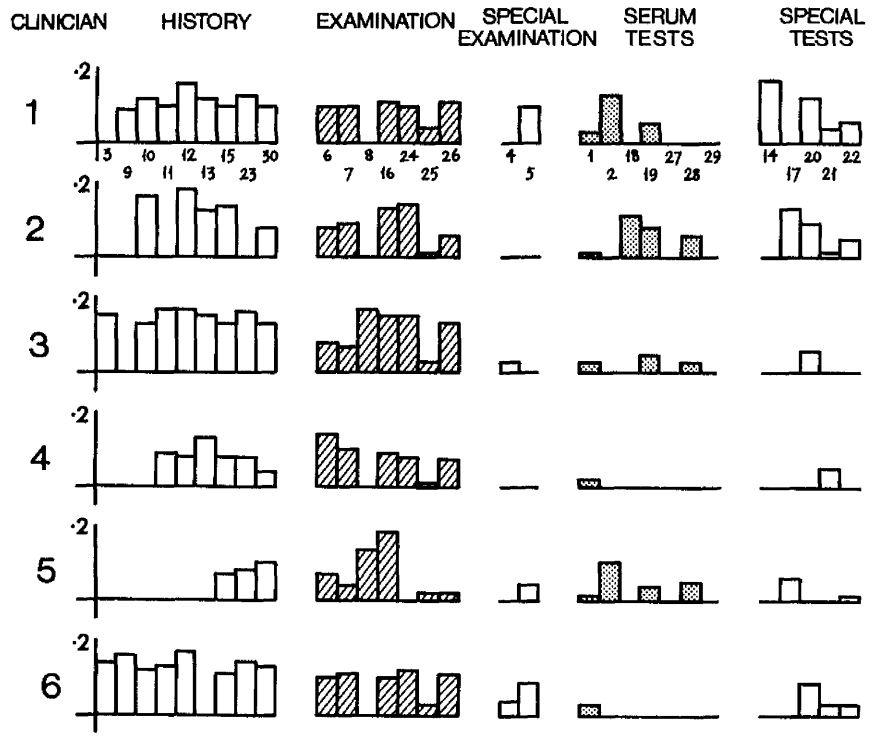
"CONSERVATIVISM/LIBERALISM"

	Top Two-Liberalism				Bottom Two - Conservativism			
	T	F	T	F	T	F	T	F
1	30	20	20	11	25	13	2	2
2	15	3	13	2	19	1	25	20
CLINICIAN 3	8	2	28	1	20	1	25	20
4	26	12	6	4	25	14	30	20
5	17	1	8	3	16	1	28	1
6	30	20	24	18	4	1	27	3

T = NUMBER OF THE TEST

F = FREQUENCY OF TEST SELECTION

FIG. 5.22. RELATION OF MEASURES OF DISCREPANCY TO FREQUENCY OF TEST SELECTION.



HISTORY:
3 Discomfort
9 Pain in goitre
10 Hoarseness
11 Dysphagia
12 Choking or tightness
13 Cough or stridor
15 Recent increase in size
23 Duration
30 Age

EXAMINATION
6 Fixation to tissues
7 Cervical lymph glands
8 Pyramidal lobe
16 Nodular or diffuse
24 Estimated size of gland
25 Consistency
26 Clinical status

SPECIAL EXAMINATION
4 Tracheal deviation
5 Laryngeal palsy

SERUM TESTS
1 Precipitine test
2 Serum globulins
18 Gamma globulins
19 ESR
27 CF Test
28 Thymol turbidity
29 Zinc sulphate turbidity

SPECIAL TESTS
14 $KClO_4$ discharge
17 BE ^{131}I
20 24 hour Thyroidal ^{131}I
21 PB ^{127}I
22 48 hour PB ^{131}I

FIG. 5.23. OVERALL PATTERN OF TEST SELECTION.

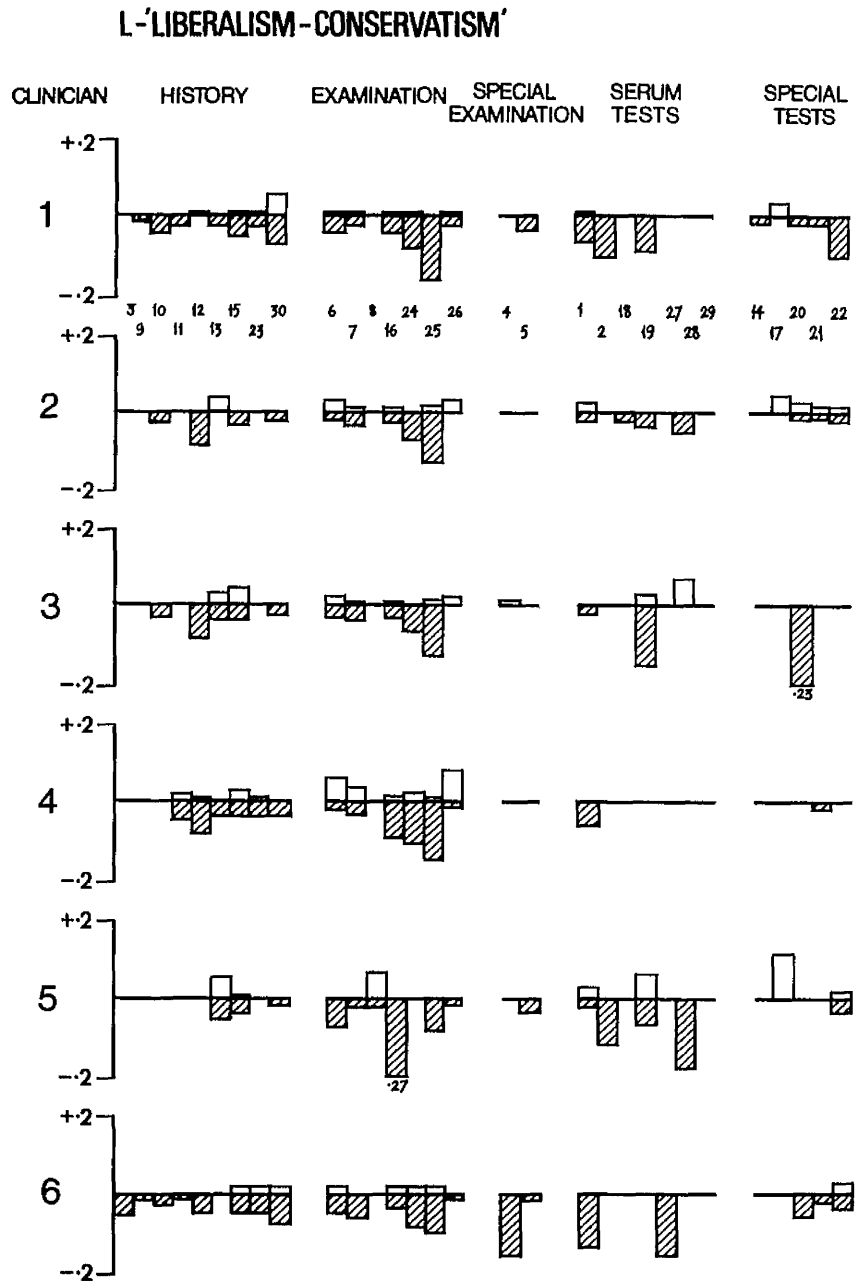


FIG. 5.24. OVERALL PATTERN FOR CONSERVATIVISM/LIBERALISM.

In the case of the conservativism/liberalism measure no clear conclusions can be drawn.

The overall pattern of test selection discrepancies is shown for all 20 cases in Fig. 5.23. In general, history items are poorest, with clinicians 4 and 5 making best use of the history. In the laboratory tests the serum tests are not selected effectively, especially by clinician 3.

A similar analysis of the conservativism/liberalism measure (Fig. 5.24) shows a general tendency to undervalue the physical examination data with clinicians 1 and 6 being the least liberal overall. The consistency of the thyroid gland (25) appears to be the most undervalued test of all. This is because it is the most useful test of all but, although clinicians value it highly, they still fail to appreciate how useful it is in this group of diseases.

5.8.6 Personality factors

In the course of the study the influence of personality factors was considered as a possible explanation of the wide individual variations in diagnostic strategies.

It/

It was decided to use two well-established questionnaires - the Eysenck personality inventory (Eysenck and Eysenck, 1965), and the 16 personality factor questionnaire (Catell, 1963), to give an overall assessment of personality factors. Because of the large amount of information provided by the '16 personality factor' test, it was decided to concentrate on obsessionalism which, intuitively, seemed the most promising starting point.

When obsessionalism appeared to be of some significance a further questionnaire specifically concerned with obsessionalism was used. This questionnaire (Ingram, 1970) isolates two second order factors f_1 - 'orderliness', and f_2 - 'indecisiveness', and is based on a population of 300 normals. All questionnaires were scored independently by an experienced clinical psychologist. Scores for extroversion were derived from the Eysenck personality inventory but with the '16 personality factor' test the clinicians were merely ranked for their degree of obsessionalism. This ranking was, in fact, identical to the overall ranking for obsessionalism derived from the obsessionalism questionnaire before the scores for the two second order factors were calculated. The results of the obsessionalism ranking/

OBSESSIONALISM (16 P.F.)	EXTRAVERSION (E.P.1)	OVERALL OBSESSIONALISM (Ingram)
1	3	1
6	2	6
5	4	5
4	5	4
2	6	2
3	1	3

FIG. 5.25. RANKING OF VALUES IN PERSONALITY TESTS.

OBSESSIONALISM (16 P.F.)	EXTRAVERSION (E.P.I)	'ACCURACY'	TOTAL TESTS
1	3	1	1
6	2	6	6
5	4	5	3
4	5	4	5
2	6	2	4
3	1	3	2

FIG. 5.26. RELATIONSHIP BETWEEN PERSONALITY FACTORS, ACCURACY AND NUMBER OF TESTS.

OBSESSIONALISM SCORES FOR CLINICIANS IN STUDY

		1	2	3	4	5	6
F_1	'orderliness'	9	6	7	3	14	12
F_2	'indecisiveness'	3	5	4	3	10	7

FIG. 5.27. SECOND ORDER FACTORS FOR OBSESSIONALISM.

AVERAGE VALUES FOR THE 'PROFILE' MEASURES

Clinician	Test Selection	Inference	Liberalism	Conservativism	Average 'U' per test	Average 'U' before diagnosis	F_1	F_2
1	.0997	.0199	.0057	.0425	.0156	.2433	9	3
2	.0930	.0461	.0142	.0422	.0117	.1279	6	5
3	.1127	.1971	.0228	.0542	.0178	.1213	7	4
4	.0787	.1197	.0206	.0534	.0117	.0641	3	3
5	.0669	.0733	.0210	.0580	.0074	.0289	14	10
6	.0972	.0416	.0015	.0532	.0156	.2251	12	7

FIG. 5.28. MEAN VALUES FOR PROFILE AND PERSONALITY MEASURES.

ranking from the '16 personality factor' test and the extroversion ranking from the Eysenck personality inventory are shown in Fig. 5.25. This shows that these two rankings are the exact converse of one another, which is normal.

When the total number of tests used, the diagnostic accuracy and obsessionalism are reviewed some association is apparent (Fig. 5.26). No obvious relationship was found with test selection, inference or conservativism/liberalism values.

The values for the two second order factors for obsessionalism are shown in Fig. 5.27. There was no clear relationship with any of the profile measures by inspection but there was a correlation coefficient of .81 between conservativism and f_2 - 'indecisiveness' which was almost significant at the .05 level. Factor f_2 also correlated negatively with 'average uncertainty before diagnosis' (5.8.3) but its correlation did not reach significance (Fig. 5.28). There were no obvious correlations between any of the profile measures and f_1 - 'orderliness'.

5.8.7 Clinical experience

In/

THYROID EXPERIENCE (YEARS)	GENERAL CLINICAL EXPERIENCE (YEARS)	AVERAGE UNCERTAINTY PER TEST	TEST SELECTION
5	5	5 (.0074)	5(.0669)
2	4	4(.0117)	4(.0787)
6	2	2(.0117)	2(.0930)
1	6	6(.0156)	6(.0972)
4	1	1(.0156)	1(.0997)
3	3	3(.0178)	3(.1127)

Fig. 5.29. RELATIONSHIP OF MEAN PROFILE MEASURES AND CLINICAL EXPERIENCE.

In any realistic attempt to assess diagnostic performance some account must be taken of the level of clinical experience of the clinicians. In this study clinical experience (defined as the number of years of daily contact with practical clinical medicine) was estimated independently by a colleague familiar with all the clinicians. The clinicians were ranked for both general clinical experience and specialist thyroid experience. This assessment did not correspond to age or to years since graduation, since individual clinicians had spent varying lengths of time in research and other non-clinical activities.

The ranking for general and thyroid experience is shown in Fig. 5.29, with the ranking for average test selection measure per test and for 'average uncertainty per test' (5.8.3). The ranking for the latter two measures is identical to that for general clinical experience. No obvious relationship with thyroid experience was found.

These results may suggest that test selection is a basic diagnostic skill which improves with general rather than specialist experience. The latter association between average uncertainty per test and general experience suggests that more mature clinicians move more deliberately towards a diagnosis, spending less time in the centre of the triangle.

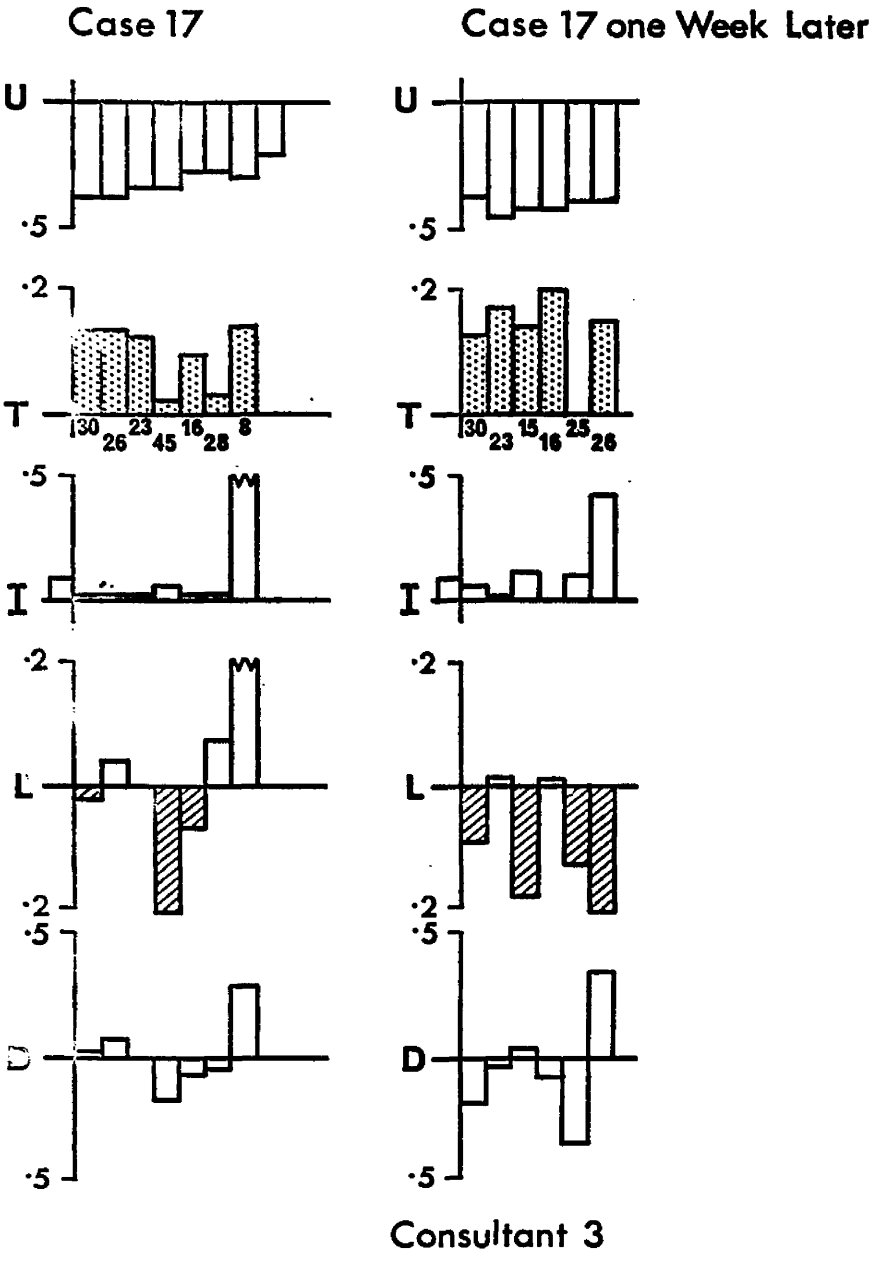


FIG. 5.30. EXAMPLE OF CASE PROFILE WITH HIGH DIAGNOSIS VALUE (> .15) WITH HIGH LIBERALISM (> .10).

5.8.8 'Channel capacity' for information

During the analysis of the clinicians' performance the observation was made that many cases were diagnosed after about 7 tests (even in cases where the computer used more than 7).

The number of tests taken by each clinician in each of the 20 cases is shown in Fig. 5.14; these results are summarised in Fig. 5.15. The individual peaks shown in this figure may be related to the capacity of individual clinicians for accumulating information before making a final diagnostic decision.

The gathering of clinical data from a patient by a clinician can be viewed as an exchange of information across a communication 'channel'. It may well be that the peaks shown in Fig. 5.15 represent the 'channel capacity' of individual clinicians for information handling.

It is therefore of interest to note that Miller (1956) published a famous paper entitled "The magical number seven, plus or minus two: some limits on our capacity for processing information". He showed, as an experimental psychologist, from a survey of his own and other research in this area that the span of absolute judgement is limited/

limited to a channel capacity of about seven plus or minus two - whether the problem is to identify auditory tones, levels of loudness or taste intensities. Similarly, he noted that the span of immediate memory for many different kinds of test patterns was about seven items in length. It may well be that the number of discrete items of clinical information (test outcomes) that a clinician can process is limited in a similar way.

5.8.9 'Pattern recognition'

An important element of the diagnostic process is the phenomenon of pattern recognition. This is usually thought of as a predominantly visual process. The same term is, however, often applied to the rapid recognition of a diagnosis using relatively little information. This latter type may depend on the intuitive recognition of some complex interdependence between test outcomes allowing the clinician to make very rapid and efficient use of clinical information.

This phenomenon may have occurred during a test case in this study and one possible method of detecting it would be to use the 'liberalism' measure. If the value for this measure is high it means that the clinician is over-estimating the value of a test outcome. In some instances this/

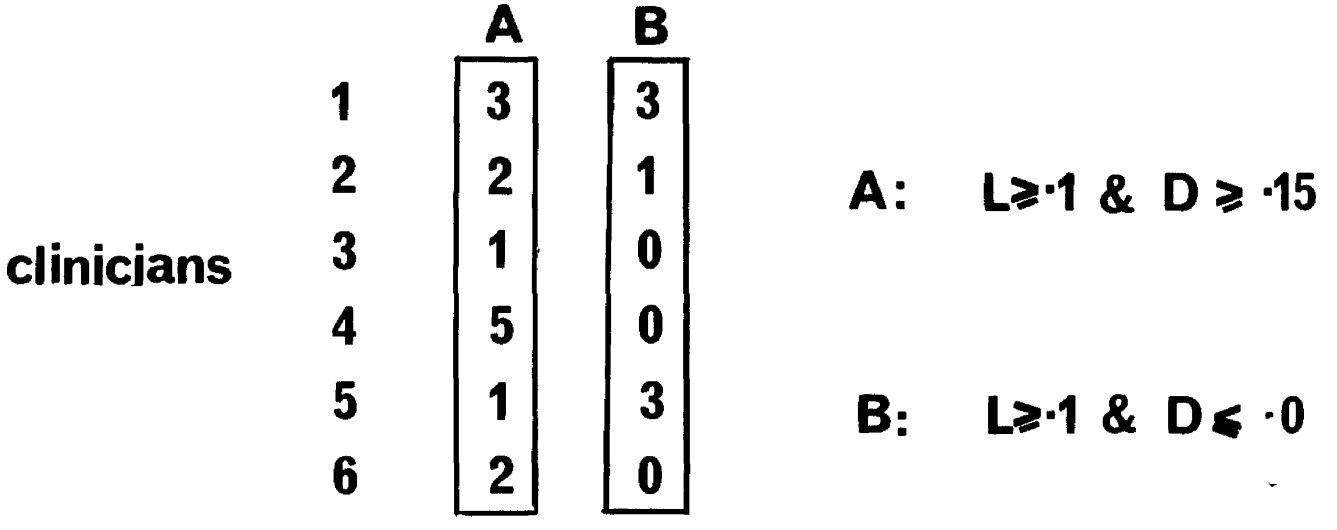


FIG. 5.31. EPISODES OF LIBERALISM.

this may be due to error or to guessing, but in others it may represent to the type of pattern recognition just described. The computer model operates on an assumption of independence between tests and may miss the pattern of interdependence which the clinician uses intuitively.

By combining the liberalism measure with the headway measure it may be possible to separate guessing from pattern recognition. Thus, if the high liberalism value is always in the direction of the correct diagnosis (i.e. with a high 'diagnosis' value) then pattern recognition may be occurring. If the 'diagnosis' value is negative then the information from the test outcome has been misestimated in the direction of a wrong diagnosis.

In an attempt to separate these two types of high liberalism values, the 20 cases were analysed for instances of high liberalism ($>.10$). These were divided into those with a high diagnosis value (Fig. 5.30) ($>.15$) and those with a negative diagnosis value ($<.0$). The results shown in Fig. 5.31 show that only clinician 3 consistently showed the pattern recognition type of liberalism value; clinicians 1 and 5 were as often wrong as right.

5.8.10 Financial costs of tests

In the design of the study, attempts were made to minimise the effects of financial and other costs. Clinicians were constantly reminded that all tests were equally available and without financial cost.

However, as an adjunct to the main study, the clinicians were asked to estimate the financial cost (to the nearest shilling) of the laboratory tests used in the study. The results of these estimates are shown in Fig. 5.32. A calculation of the actual financial costs is included, based on a detailed survey (by the Department of Operational Research at Strathclyde University, Glasgow) of the individual laboratories involved.

The wide discrepancies in clinicians' estimates of costs confirm how little insight the clinicians have into the financial costs of the investigations which they are using every week in their routine practice.

5.9 Discussion

5.9.1 Introduction

The aim of this study is to develop techniques for the measurement of individual differences in the diagnostic strategies/

	clinician						
	1	2	3	4	5	6	s.
Precipitin test	120	10	5	12	5	30	3.0
Serum globulin	40	15	4	4	20	10	2.0
G-globulin	40	25	4	5	20	50	12.0
E.S.R.	20	8	1	1	5	5	2.0
$ZnSO_4$ turbidity	40	20	4	2	20	50	5.0
PBI^{127}	300	25	7	20	20	50	5.0
$BE.I^{131}$	40	20	40	5	40	10	8.2
$KCLO_4$ discharge	240	-	60	20	20	70	20.0
24 I^{131} uptake	200	10	50	5	40	30	1.0
48 PBI^{131}	140	13	50	5	20	30	7.0
T_3 Resin sponge	200	17	16	15	20	10	11.0
T_4 Suppression test	600	30	70	40	100	90	15.0
Thyroid scan	400	30	80	5	100	50	-
TRL. Agglut.	120	15	5	12	20	70	10.0
Thyr. Microsomal Antibody	120	20	5	10	40	90	
TSH Stimulation test	600	200	70	300	300	90	143.0

FIG.5.32. CLINICIAN'S ESTIMATE OF COSTS WITH ACTUAL COSTS FROM SURVEY.

strategies of clinicians. The study also represents a contribution to the long-term process of 'mapping-out' the interface in clinical decision-making between clinician and computer.

Clinicians are studied in detail while tackling rather formalised decision problems in order to identify, (i) those decisions which are most effectively made by computers, (ii) those in which clinicians excel and, finally, (iii) those decisions where the combination of clinician and computer offer the best approach to a solution.

Before the widespread availability of computers, the objective assessment of personality by clinical psychologists was an important focus for the debate about the relative merits of clinical and statistical personality assessment. The most important early statement of the problem was in the book 'Clinical versus Statistical Prediction' published in 1954 by Professor Paul E. Meehl. In this book, which was concerned primarily with prediction techniques in clinical psychology, the problem was posed as a choice between methods of decision-making which are based on the clinician's rational and informed reflection on the available/

available information, and those methods which are dependent on a statistician or a computer program.

Meehl's book brought together and presented objectively arguments which had been developed intermittently in psychological circles for some years. He was, if anything, inclined to favour the clinician. Many studies in clinical psychology have followed this theme in recent years without a clear resolution of the problem. The advent of computers has tended to focus more attention on the problem. Much of this recent research has been critically reviewed by Sines (1970).

Many of the original studies by Meehl (1946, 1960) were based on comparisons between statistical methods and the clinical judgement of clinical psychologists in scoring a well established objective personality test - the Minnesota multi-phasic personality inventory (M.M.P.I.).

Kleinmuntz (1963) has continued such comparative studies on the MMPI but has more recently (Kleinmuntz, 1968) studied the diagnostic behaviour of a group of clinical neurologists. In these latter studies he decided to concentrate on neurology because of the highly structured nature of the clinical data available. The

experimental/

experimental design was based on the well-known parlour game "Twenty Questions". The examiner chooses a disease which the clinician has to identify by asking questions to which there are Yes/No type answers.

In essence, the clinician chooses a route to a diagnosis along a logical decision tree (2.2). These decision trees are then analysed in detail. Kleinmuntz (1968) found that diagnostic accuracy was greater and questions were fewer among the more experienced clinicians. Such clinicians are less distractible (by anomalous results) and more consistent than their comparatively inexperienced colleagues. A very similar study was conducted by Wortman (1965) on a small group of three experienced neurologists. This study also produced logical decision trees and the conclusions were almost identical to those of Kleinmuntz (1968). Both of these authors found that the more experienced clinicians selected the more informative tests. This finding corresponds to the association between test selection and general clinical experience described above (5.8.7). However, in these studies the informativeness of tests was not calculated formally (as in the present study) but was reached empirically by agreement with other experienced neurologists.

A similar approach to the analysis of diagnostic performance is that of Rimoldi and his co-workers in the Psychometric Laboratory of Loyola University (Rimoldi, 1961). This group developed a 'test of diagnostic skills' with the aim of studying how a medical student proceeds to a diagnosis.

The test uses a set of cards, each with an item of information printed on one side, which were grouped under such headings as 'interview and history' 'physical examination' and 'laboratory procedures'. The student asks for an individual card, reads the data on the back of each, then asks for another card and proceeds in this sequential fashion until he is satisfied with the diagnosis.

The analysis of test performance is based on such features as accuracy of diagnosis, number of cards requested, sequence of cards and the 'value' of cards chosen. This last score is based on a comparison with experienced clinicians.

The most important finding with this technique (which is mainly used for teaching purposes) is that the overall level of performance rises with clinical experience.

Apart/

Apart from these two groups of studies there has been no systematic attempt to study the diagnostic behaviour of clinicians in any detail.

In the study described in this chapter the techniques used have much more in common with the work of Phillips and Edwards (1966) on statistical studies of human decision-making. These authors compared human decision-making behaviour with a statistical model (using Bayes' theorem). They have concentrated on testing individuals as probability estimators in experimental game playing situations where small monetary rewards are paid for correct decisions. In all of their studies they have shown that human beings are conservative probability estimators(Edwards, 1968).

The present study using only three diseases has the advantage of retaining the simplicity of these artificial experiments while being accepted as reasonably realistic by the clinicians taking part.

This study differs from the above studies in several ways:

1. It is based on a sequential probabilistic model.

2./

2. It concentrates on the information processing aspect of the diagnostic process
3. It uses formal measurements of the informativeness of tests and of test selection.
4. It measures liberalism as well as conservativism in probability estimations.
5. It attempts to characterise diagnostic strategies in a formal way by such measures as 'average uncertainty per test' and 'average uncertainty before diagnosis'.
6. It includes some attempt to demonstrate the influence of personality factors on diagnostic performance.
7. It attempts to detect and measure one type of pattern recognition in individual clinicians by means of the liberalism and headway measures.

The results of the study will be discussed under the appropriate headings. An important general reservation about the interpretation of the results in the discussion which follows is that the sample of six clinicians is small and so no firm conclusions can be drawn. The study itself is seen as an exploratory one, concerned with the development of techniques and concepts which will be used in larger studies in the near future.

5.9.2/

5.9.2 'Diagnostic accuracy'

The exact significance of the range of diagnostic accuracy shown in Fig. 5.12 is difficult to assess. The range is narrow but seems to bear some relationship to overall obsessionalism. However, when viewed in relation to the number of tests used, it seems likely that clinician 1 and 6 in particular persist in collecting data until they are absolutely certain of a diagnosis. By contrast, clinician 2 is perhaps more impulsive in hazarding a diagnosis with many fewer tests. Clinician 3 who had most misdiagnoses was the least experienced. Clearly, a compromise must be reached between over-investigation and inaccuracy. This will be especially important if either the tests are expensive or unpleasant or if the patient's condition is serious.

5.9.3 Number of tests used

The influence of obsessionalism on the number of tests used has already been noted, while clinical experience seems to have no obvious influence. The importance of balancing the number of tests used against diagnostic accuracy has also been emphasised, but if personality factors are more important than experience then the possibility of reducing the number of tests used by/

by clinicians is not strong. For example, clinician 1 takes almost double (213) the number of tests as clinician 2 (120). It may well be that a computer-based selection of tests is the only way of circumventing the influence of obsessionalism and reducing the number of tests needed to make an accurate diagnosis.

In the analysis of 'reproducibility' or 'consistency' the influence of clinical experience is not clear (Fig.5.16). Clinician 5 is much the most consistent as well as being the most experienced, but he is probably in any case the most effective clinician in the group when all measures of performance are taken into account.

5.9.4. Individual routes to a diagnosis

The main difficulty about this aspect of the study lies in comparing such complex measures of diagnostic behaviour with one another. However, some comparisons are easier than others and there can be little doubt that clinician 5 differs markedly from clinician 2 in cases 11A and 11B. Clinician 5 is clearly much more certain of the diagnosis, yet both are very experienced clinically and do not differ greatly in their obsessionalism.

The measure of 'average uncertainty per test' shows that/

that clinician 2 consistently spends more time in the centre of the triangle than clinician 5. This measure has been shown to bear some relationship to general clinical experience (Fig. 5.29) and it may be that the 'deliberateness' with which a clinician moves to a diagnosis (which is reflected in this measure) is governed by some underlying strategic diagnostic skill which improves with experience.

The 'average uncertainty before diagnosis' is a measure of how near a clinician moves to a diagnosis before actually choosing it. It is negatively correlated with f_2 - indecisiveness, and suggests that the more obsessionally indecisive clinicians move very close to a diagnosis before feeling certain enough to choose it.

These formal empirical measures can only go a little way to characterising the individuality of routes taken by clinicians. A great deal of information about individual strategies remains unmeasured. It is possible, however, by inspection to note individual diagnostic styles. Thus, clinician 6 tends to accumulate information at certain stages in his route before moving off towards a corner (Fig. 5.9). Clinician 2 appears in case 13 (Fig. 5.19) to be very indecisive but this is in large part due to the atypical/

atypical features of this case. Yet clinician 6, normally indecisive, deals with this same case very much more effectively than clinician 2.

In future studies such cases, where data is atypical or incomplete, will be used to 'stress' the diagnostic abilities of clinicians and to highlight individual differences in diagnostic styles.

5.9.5 'Profile analysis'

By contrast with the data on individual routes, the 'profile' data is much easier to analyse. An important use of this type of data is the overall view it gives of the use of history, physical examination and other data. It helps to highlight particular tests which individuals or groups undervalue or select ineffectively and this feature is used extensively in the teaching system described in the next chapter.

Test selection shows an association with general clinical experience but not specialist thyroid experience (Fig. 5.29). This suggests that the ability to select investigations effectively is a skill basic to years of clinical experience. The lack of association with thyroid experience may mean that this skill can be transferred to other groups of diseases more easily in experienced clinicians.

If this skill can be learnt (as the association with clinical experience suggests) then it may be possible to teach undergraduates or postgraduates to select tests more effectively (for example, by the computer system described in the next chapter).

By contrast, it may not be so easy to teach clinicians to be less conservative since this is not related to clinical experience but to f_2 indecisiveness, which may be difficult to modify. In support of the argument that these skills may not be amenable to teaching is the fact that clinician 3, who is by far the most consistently liberal (Fig. 5.30, Fig. 5.31), is the least experienced of the group.

The use of the liberalism measure in detecting pattern recognition is discussed later.

5.9.6 Personality factors

The influence of the personality factors examined in this study has already been discussed in relation to diagnostic accuracy, number of tests, individual routes and the diagnostic profile. Since these factors seem to influence the intellectual, information processing aspects of the diagnostic process it seems likely that specific personality/

personality characteristics may bring with them specific cognitive styles which come into play in any form of decision-making including diagnosis. For example, obsessionalism may influence individuals tolerance for the uncertainty involved in diagnosis. The more obsessional clinician can be seen to take many more tests, to assess them more conservatively but be more persistent in being absolutely certain about a diagnosis. Some clinicians may act impulsively to reduce the level of uncertainty, while others clearly react to uncertainty by collecting more information but end up being more uncertain at the end than they were at the beginning.

It is surprising that f_1 'orderliness' is not shown to have an influence but this may be due to the size of the sample and the fact that most cases in the study were uncomplicated. It is likely that this factor would be of greater importance in complex cases where the orderly assessment of a large of complex set of tests would be an advantage.

However, the major problem associated with the influence of personality factors is that they are unlikely to be amenable to change even in the medical student. While this observation would have to be verified in later experiments/

experiments it seems likely that the main value of demonstrating any association between personality and diagnostic styles would be in self-selection of medical specialities. Thus, obsessionalism if marked may be a considerable advantage (especially if there is a preponderance of f_1 'orderliness') in research, but a positive disadvantage in some areas of clinical medicine where rapid flexible decision-making is in great demand.

It is possible that by providing insight into the influence of such factors that individuals, early in their careers, may select areas of medicine appropriate to their styles of decision-making. Failing this, they may gain enough insight into the effect of their type of personality to avoid certain undesirable traits, e.g. impulsive diagnosis to which they are particularly prone.

5.9.7 Clinical experience

The possible influence of general clinical experience of test selection and on the 'deliberateness' of an individual route to a diagnosis (as measured by the average uncertainty per test) have already been discussed.

One of the most important aspects of future studies will be to pursue in more detail the influence of experience and/

and to devise methods, such as computer-assisted instruction, for promoting the more rapid acquisition of diagnostic skills. Clearly if the exact relationship of particular types of clinical experience to diagnostic skills can be studied, the influence on undergraduate and postgraduate teaching will be considerable.

It is hoped that specific skills, such as test selection, can be isolated and studied. It seems likely that what is known as 'diagnostic acumen' in fact consists of a group of skills, differing from clinician to clinician in their state of development. Much of what is normally referred to as 'observer error' must in fact be due to such individual differences in clinical skills.

It is likely that it will be in the more complex cases with atypical features that the experienced clinician will excel. It will be important in future studies to try to confirm the findings of Kleinmuntz (1968) that experienced clinicians are less likely to be less easily distracted from a diagnosis by anomalous information appearing during the assessment of a case.

5.9.8 'Channel capacity'

The finding of peaks in the number of tests used by clinicians/

clinicians and the possible relationship to 'channel capacity' has been suggested earlier (5.8.8.).

If these peaks do represent the capacity of clinicians for optimal processing of data then clearly this is important in day to day decision-making. It would therefore be counter-productive if these findings are confirmed to present a clinician with 12 to 20 bio-chemistry findings from an auto-analyser at one time.

The original idea of the 'magical number 7' and of channel capacity (Miller, 1956) has been criticised as being too passive an idea of information handling by the brain. It is now thought that the process of detecting and processing information is a much more dynamic and active one. This does not, however, invalidate the suggestion that individuals have a capacity for optimal processing of information and that this is frequently exceeded.

It may well be that the number of tests used in a case is influenced by the short-term memory capacity of the clinician who must retain all the data for the final diagnosis. It would be interesting in future studies to compare the short-term memory capacity of, for example, clinicians/

clinicians 1 and 6 with the others in the group to see if their larger total number of tests is related to a better capacity to retain information.

5.9.9 'Pattern recognition'

The association between 'liberalism' and 'diagnosis' described in 5.8.9. and its possible relationship to 'pattern recognition' and 'guessing' is, of course, purely speculative. However, the fact that successful liberalism is so predominant in one clinician (No. 3) who is also the least experienced, suggests that it may be worth pursuing in later studies. If it does represent a more effficient method of information handling, it may well be inate rather than acquired by experience.

The particular pattern of tests which precede these 'liberal surges' appears to be common to all such instances but in fact is typical of almost all the 20 cases completed by clinician 3. However, future studies would involve the 'spotting' of possible 'pattern recognition' among clinicians. By this type of analysis they could then be studied in much more detail.

5.9.10 Financial costs

The wide discrepancies between clinicians estimates of/

of costs and the actual values has been shown in Fig. 5.32. It is clear that they have little idea of costs and hence it is likely that they play relatively little part at present in the selection of tests by the clinicians in this study.

5.9.11. 'On-line profile analysis'

During the analysis of the study it was decided to develop a series of programmes to provide the profile analysis almost instantaneously at a computer terminal. The techniques and examples of the output are described in detail in the next chapter.

It is intended that these techniques will be used not only for teaching but also to provide a 'test-bed' for the detailed, flexible, analysis of decision-making by individual clinicians (such as those suspected of using pattern recognition). It will also be used to investigate situations where only the combination of the particular skills of the clinician and the power of the computer can provide solutions to diagnostic problems.

C H A P T E R 6 : THE ANALYSIS AND TEACHING OF DIAGNOSTIC SKILLS

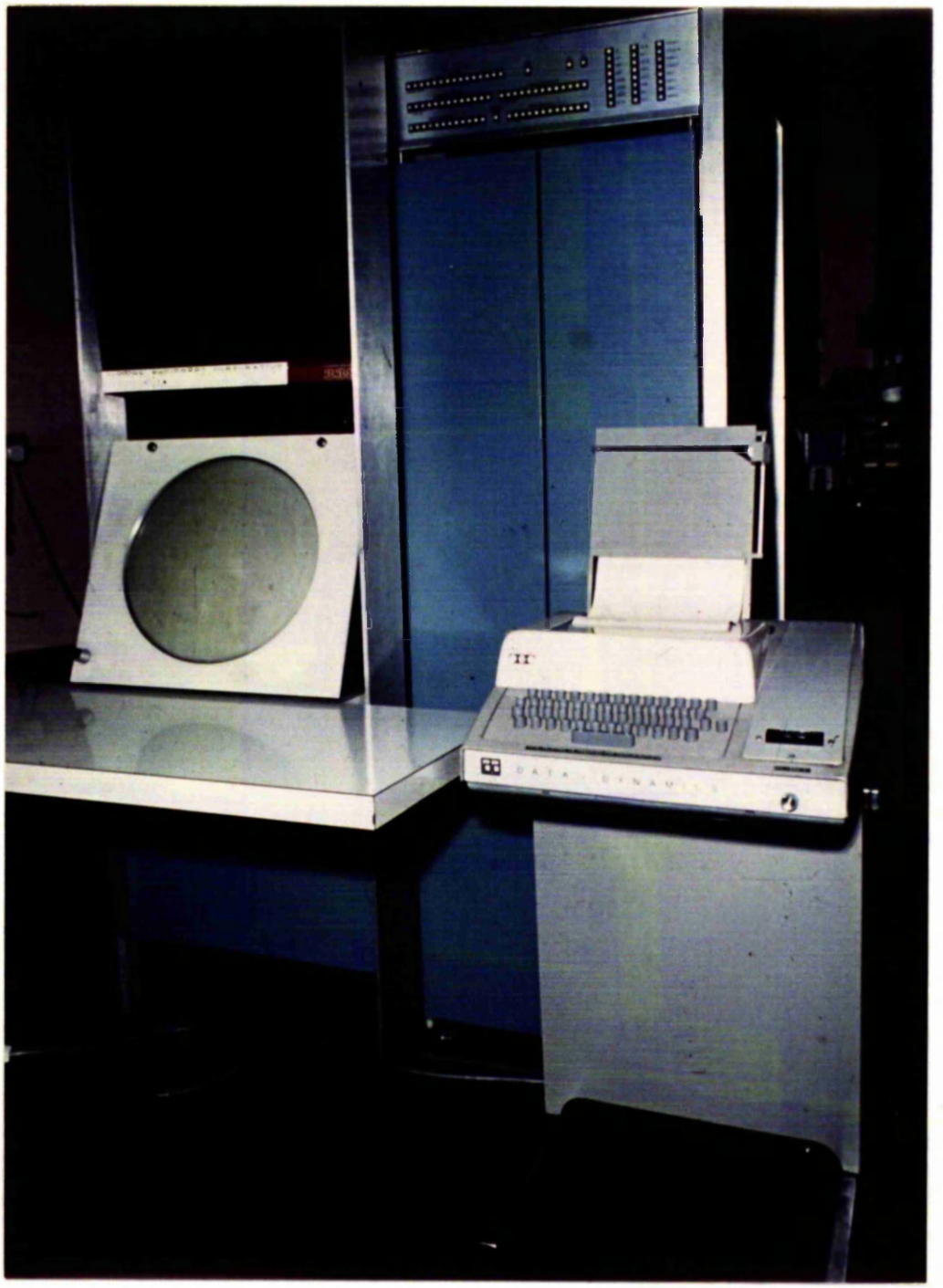


FIG. 6.1. INTERACTIVE VISUAL DISPLAY (PDP8/338) WITH TELETYPE.

6.0 Summary

The technique of 'profile analysis' which was developed to analyse the diagnostic performance of the six clinicians in the study of decision makers has been adapted to analyse and teach some of the skills involved in the diagnostic process.

The computer program used in this technique has an accuracy comparable to an experienced clinician. By providing the profile analysis at a computer terminal after each decision is made it is hoped that the student can be taught to perform as well as the computer.

A standard teletype was found to be too cumbersome, too slow and too noisy as a terminal for using the technique. An interactive visual display with a light pen was found to simplify communication between student and computer. It allowed rapid entry and display of data. Complex relationships between the elements of the profile were displayed without any numbers by means of histograms.

A system with four recurrent 'frames' of data display with their associated responses from the student has been developed and is now available for trial with students.

The place of such a system in conventional medical education/

education, in educational technology and among the available systems of computer-assisted diagnosis is briefly discussed.

6.1 Introduction

In the last chapter a technique for the analysis of the diagnostic performance of clinicians was presented. The main component of this analysis was the diagnostic profile whose theoretical basis has already been described in detail (5.6).

The profile is based on discrepancies between the clinician's performance and that of the computer program. If the student (undergraduate or postgraduate) could be taught to reduce these discrepancies to nil, then the student's accuracy of diagnosis would be the same as that of the computer. Since the accuracy of the computer program is 93 per cent (4.4) which is comparable (in this set of diseases) to an experienced endocrinologist, then there exists a basis for a teaching system.

The teaching system can function at a computer terminal by providing the profile analysis as soon as the student has chosen his test and has altered his estimate of the probabilities. If he is then allowed to reconsider both his/

his test selection and his probability estimate in the light of this analysis he is then in a position to learn.

6.2 The study

The data (likelihoods, prior probabilities and test cases) used in this study were identical to those described in Chapter 5. The 'profile analysis' program was the basis for the programs developed in this study.

The study consisted of two parts. In the first, the 'profile analysis' program was modified to allow the prior probabilities, choise of test and probability estimates to be entered at a teletype by the student while the test outcome and profile analysis results were printed out at the same teletype by the computer. In the second study, the same 'on-line' version of the profile analysis program was used but an interactive visual display was used in place of the teletype (Fig. 6.1).

Both applications used the 'on-line' time sharing system (Cotan 3) run on the I.C.L. KDF9 computer in the Computing Service Department of Glasgow University. The calculations involved in producing the diagnostic profile are large and were performed on the main KDF9 computer. The teletype is used as the input/output device in the

first/

Date ? _____

Clinician Number ?. _____

Prior Probabilities ? _____

The initial uncertainty is 0.0047

which test would you select first? _____ (30; age)

The outcome of this test is _____ ($\geq$ 60 years of age)

What are the new probabilities now? _____

Your diagnostic profile is now

* Test	U	T	I	L	D
30 (age)	.2072	.1056	.0222	.1165	.0752

Do you wish to continue (i.e. instead of trying another test in place of age) ? (Yes)

Next Test? _____ (23, duration of goitre)

The test outcome is _____ ($<$ 1 year)

What are the new probabilities now? _____

* Test	U	T	I	L	D
30 (age)	.2072	.1056	.0222	.1165	-.0752
23 (duration of goitre)	.1713	.0570	.0006	.0040	-.0027

Continue ? _____

* These figures would be displayed visually as Histograms when an interactive visual display terminal is used.

FIG. 6.2. TELETYPE PRINT OUT.

first application, being replaced in the second application by a PDP8/338 display system (Fig. 6.1).

In both studies the same test cases used in the studies described in Chapter 4 are stored on the magnetic disc file of the KDF9 computer system. They are graded into degrees of difficulty by the objective technique used in Chapter 5 (5.2).

6.3 Techniques

6.3.1 The teletype system

The student sits at the teletype and identifies himself by typing in his name and the date of the teaching session. He is asked to estimate the prior probabilities. The computer replies (fig. 6.2) by typing out the value for the initial uncertainty and asks which test he would select.

The student types in the test number and the computer 'looks up' the appropriate test outcome on the case being assessed (on its disc file) and types out this information. The student is then asked to enter the new probabilities in the light of the information provided by the test outcome.

When the student enters the probabilities the computer uses the profile analysis program to calculate the diagnostic profile which is then printed out at the terminal within a few seconds.

The student reviews the profile and is asked if he wants to continue for another cycle; to reconsider his choice of test or probabilities, or to stop and make a diagnosis.

The teaching session can continue in this manner till all available tests are used, or can stop at the end of any intermediate cycle. After a diagnosis is made the student is asked if he wants to try another case or to end the teaching session.

When this system was assessed experimentally it was found that the interchange of information by teletype between the student and computer was slow and rather noisy. The user has to familiarise himself with the teletype keyboard and the codes needed to operate it. The acquisition of these techniques tended to obscure the main purpose of the system, namely to teach certain aspects of the diagnostic process.

The main problem from the point of view of teaching is that the profile analysis is presented as a series of 4-digit numbers. This makes the scanning and assimilation of the profile information by the student rather difficult.

Because of these drawbacks it was decided that the interface between the student and the computer was much too cumbersome, noisy and slow.

In an attempt to provide an alternative interface which was unobtrusive, simple to operate and fast, the teletype was replaced by an interactive visual display.

6.3.2 The PDP8/338 interactive visual display with a light pen

The PDP8/338 display is attached to the KDF9 computer (Fig. 5.1) and information is entered into the computer by pointing to one of a number of pre-selected areas on the face of the display tube. This is done with an electronic 'light pen' which is connected to the display by a cable. (Fig. 6.3).

The display tube has embedded beneath its surface a grid of wires and with them it can locate the area on the face of the tube at which the light pen is pointed. This is possible because the light pen 'fires' a stream of electrons to complete a circuit at the point on the grid at which it is pointed. The computer program includes a record of the information displayed over each area of the tube at any given moment (such as a list of tests, Fig.6.5). In this way, information can be entered into the computer by pointing at it with the light pen.

The light pen allows the PDP8/338 display to be used as an input terminal, while output information is displayed on/

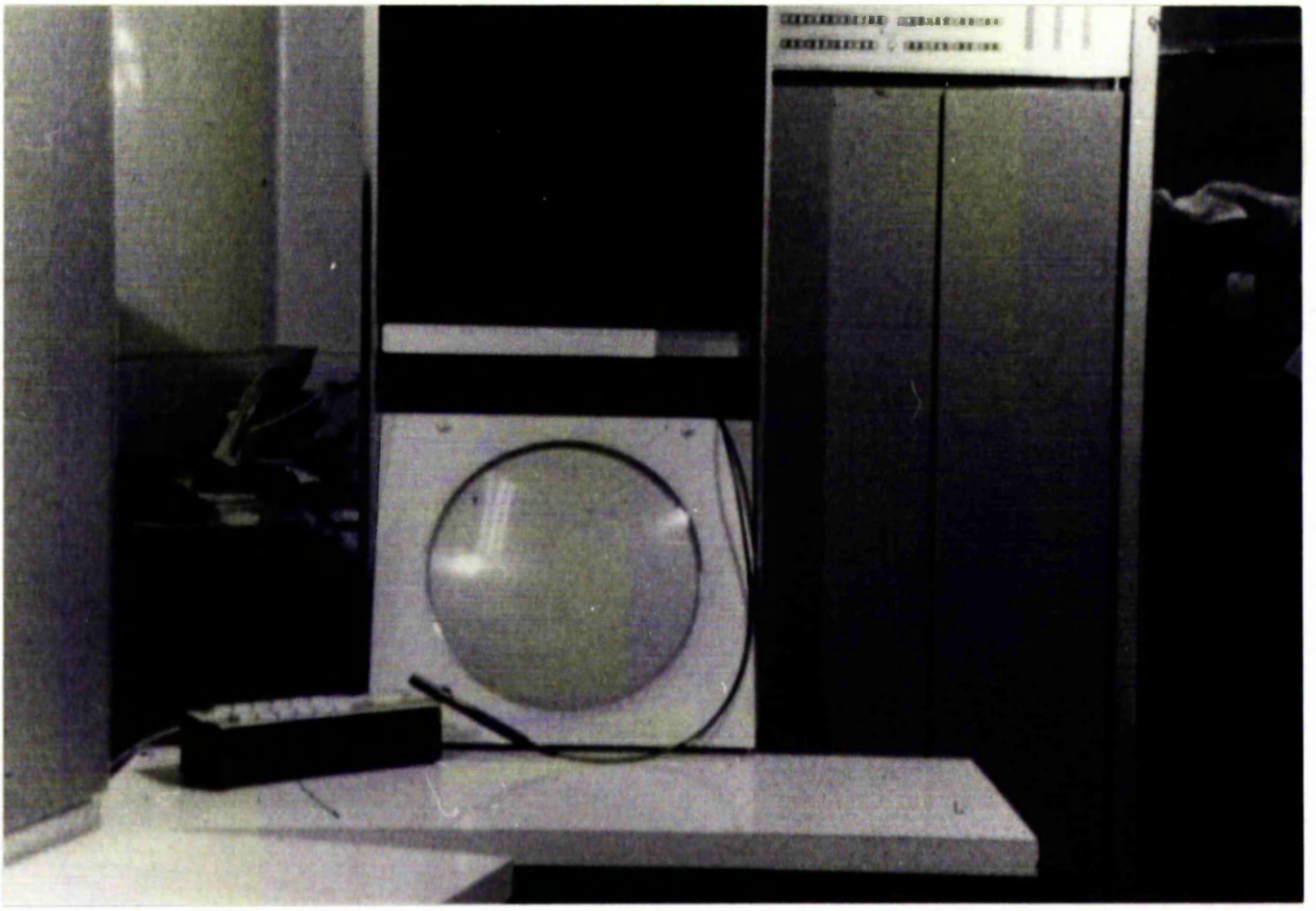


FIG. 6.3. DISPLAY WITH LIGHT PEN.

on the face of the tube, either as lines (or shapes) or as alphanumeric data. Probabilities can be entered by dividing a line (drawn on the face of the tube) into proportions with the light pen (Fig. 6.4) while the profile analysis can be displayed as a set of histograms (Fig. 6.7).

In this way the interactive visual display can provide rapid, silent interaction between student and computer. The student is then able to concentrate on the learning process without being distracted by complex numbers or by the technology involved in this type of teaching.

6.3.3 The teaching sequence

When using the interactive visual display the student goes through a teaching cycle which consists of four 'frames' of information displayed on the face of the tube.

A. A line divided into twenty units is displayed. The student indicates his estimate of the prior probabilities by dividing this line into three portions with the light pen (Fig. 6.4).

B. The second frame gives the list of available tests (Fig. 6.5). The student chooses his test by pointing at the test number with the light pen. In later cycles the list displayed has removed from it those tests already selected.

FIG. 6.4. LINE FOR INDICATING PROBABILITIES.

FIG. 6.5. LIST OF TESTS WITH LIGHT PEN.

C. In the third frame the test outcome is printed out along with a graph to remind the clinician of his choice of probabilities up to the previous cycle. After noting the test outcome and reviewing the graph (fig. 6.6) he divides up the calibrated strip in this frame to indicate his revised probabilities.

D. Within a few seconds the 'profile analysis' is displayed in the fourth frame as a set of histograms (Fig. 6.7). He is then asked if he wishes to stop, to continue to the next choice of test or to reconsider his move. If he chooses to continue, he then moves into the next cycle and the list of available tests (minus the one just chosen) is shown. If he wishes to 'backtrack' then the list appears again with his last test reinstated. He can then choose if he wishes a new test, or he may confine himself to reconsidering his estimate of the probabilities in his original choice of test.

He continues in this cyclic fashion till he is satisfied that he has reached a diagnosis. When he does so, he indicates that he wishes to stop after the 'profile' frame is shown. He will then be asked to choose a new case number or to stop. If he chooses a new case the prior probabilities are assumed to be as before, and he starts again/

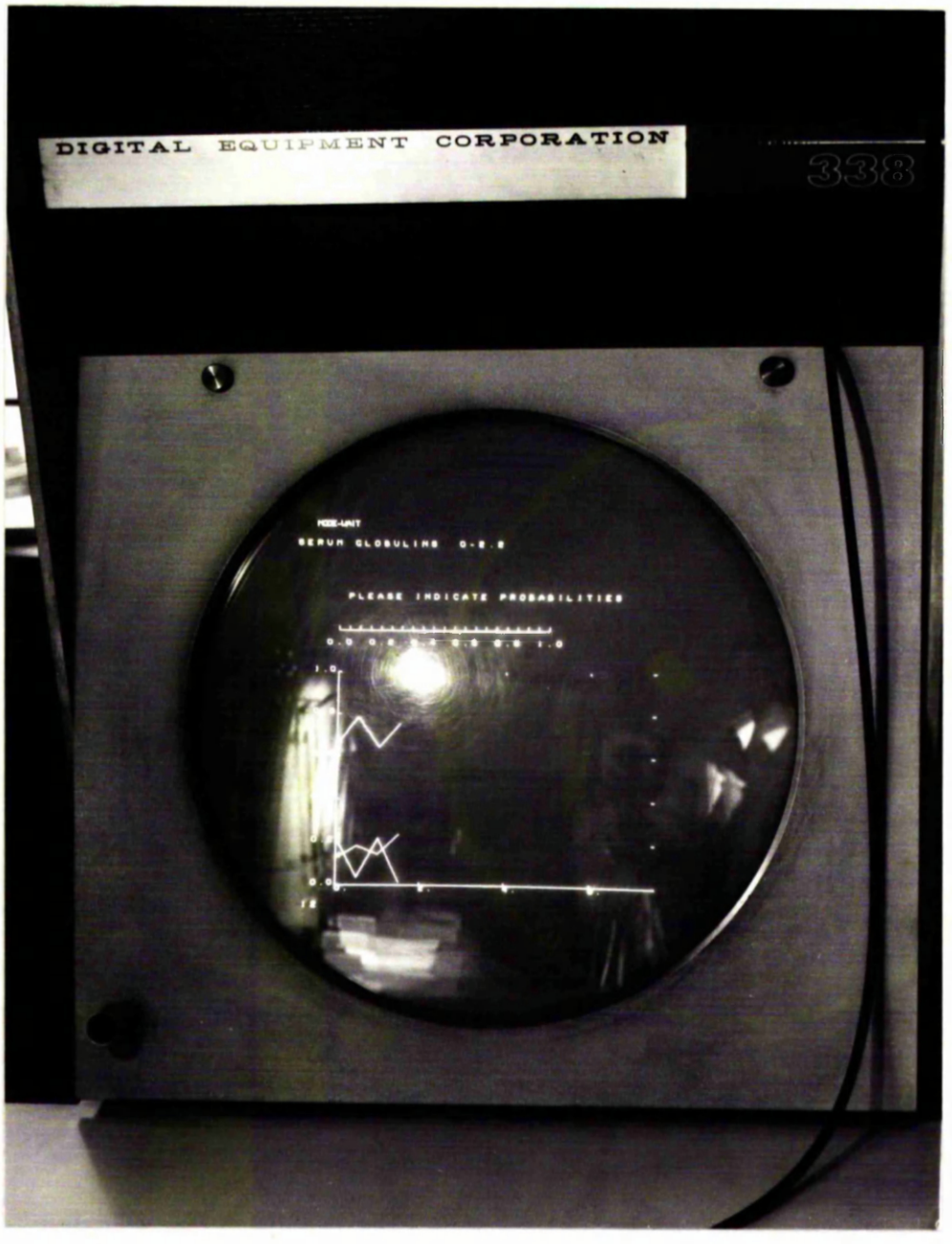


FIG. 6.6. GRAPH OF PROBABILITIES WITH TEST OUTCOME.

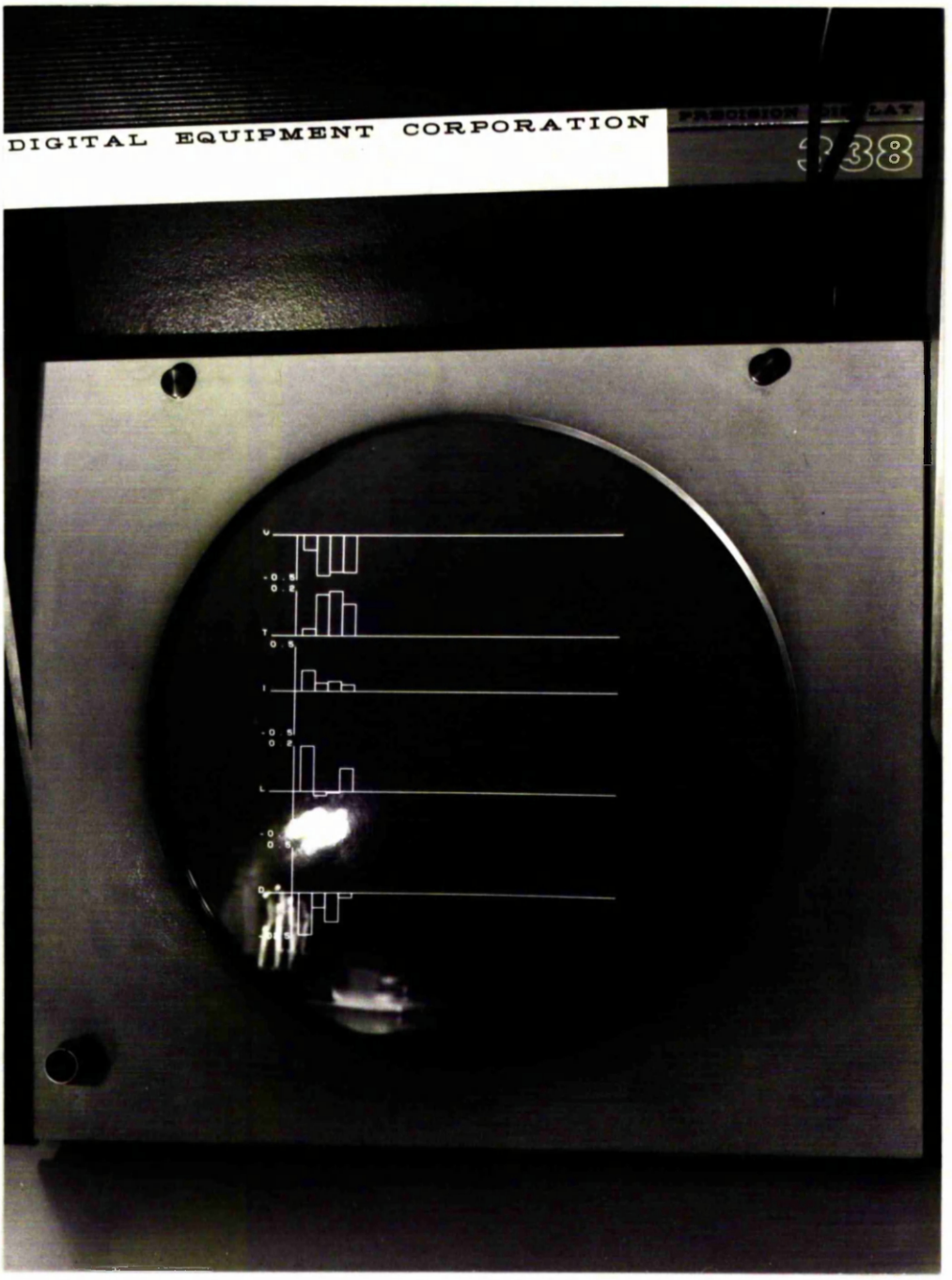


FIG. 6.7. PROFILE ANALYSIS.

again by choosing his 'best' test.

If he wishes to stop, the computer as well as recording his performance on the disc, prints out on an adjoining teletype a list of his choice of tests followed by the actual measurement of his discrepancies to four places of decimals (Fig. 6.2). It also indicates his 'backtracking'. This numerical record can be used for scoring as well as more detailed studies of diagnostic behaviour, including the diagnostic strategies of different clinicians.

6.4 Discussion

The two studies presented in this chapter deal with systems which have not been assessed by students. Discussion must therefore be confined to the place which such techniques might play in computer-assisted instruction and in medical education in general.

The field of computer-assisted diagnosis has been under development since 1958. In 1961 about five major systems existed in the United States, rising to about 100 at present (Pressman, 1970).

Most of the present day projects involve such subjects as engineering, mathematics and personnel training of many types. In medicine only a few systems exist, and none are in routine use. One of the best known examples is that of Swetz/

Swetz and Feurzeig (1965) which uses an IBM 1500 computer and an alphanumeric terminal and light pen; the system at the University of Illinois is based on a CDC 1604 computer (Pressman, 1970). In the United Kingdom the only established system is that in the "Computer-based Learning Project" at the University of Leeds (de Dombal et al, 1969, 1971). This latter system concentrates heavily on surgical gastro-enterology and used a 'Modular One' computer.

In all of these systems the basic philosophy is that of a 'Socratic' learning system based on questions and answers. The student is asked questions, gives his answer and, if wrong, is provided with varying amounts of information to remedy his deficiencies in knowledge. Such systems are the logical developments from programmed learning systems such as the tape/slide methods of teaching (Harden et al, 1969).

The system described in this study is based on information theory, decision theory and conditional probability theory. It concentrates on the information processing aspects of the diagnostic process and seems to be unique in both these respects.

It is likely that the 'profile analysis' system would be seen as a third component in an experimental automated approach/

approach to medical teaching. The first stage would use the well-established technique of tape-slide presentations (Harden et al, 1969) to present factual and conceptual information to the student. The second stage might use a 'Socratic' type of system to test out the student's grasp of the information provided by the tape/slide system. It would allow some limited analysis of the responses of individual students and provide factual and other assistance to the student in difficulties. This type of flexibility is impossible with the tape/slide approach.

The last stage is provided by the profile analysis technique developed in this chapter. This approach is even more flexible than the 'Socratic' systems. It tests the student's ability to assess clinical information with cases of varying degrees of difficulty. It emphasises the importance of selecting the best investigation and of placing the right emphasis on the value of the test outcomes. It can be also used to illustrate the relative merits of the history, physical examination and of laboratory tests.

The 'profile' analysis technique could, of course, be used alone to complement conventional teaching methods. Indeed it is likely to find a place in complementing rather than/

than replacing traditional approaches such as the lecture or clinical demonstration. It may well be of most value in teaching groups of students or in illustrating a lecture.

The system described here is an experimental one. It is possible that only part of the profile (e.g. test selection and conservativism/liberalism index) may be necessary for undergraduate teaching and that much cheaper displays can be used eventually once techniques and concepts are simplified.

In the earliest version of this system the teletype was found to be slow and cumbersome and the delay made the printing of the histograms quite impractical. The interactive display simplifies both the input and output modes. The light pen allows the student to indicate choices simply by pointing and involves no activity at the teletype. He can sit in front of the screen throughout the teaching session. The use of the 'calibrated strip' for estimation of probabilities is very simple.

The output, in the form of graphs and histograms, allows the rapid visual assimilation of highly complex information. Above all, complicated inter-relationships between the measure of performance (i.e. histograms) can be/

be demonstrated, which would be very difficult if they were displayed numerically.

The speed of interaction allows for a smooth integration of the student's response to information and the computer's analysis of his response. The 'back-tracking' facility provides great flexibility in the learning potential of any one test case. To provide such flexibility in a stage I (tape/slide) or II ('Socratic') system would be very complex indeed.

The 'profile' system has, of course, many 'game-playing' features and should provide a challenging situation for the student. The development of a scoring system which will encourage the student is being considered. The technique already allows for many different 'routes' to a final diagnosis, which would still give acceptable total scores on all the discrepancy measures which are recorded.

The concentration on the informating processing aspects of the diagnostic process is deliberate, since other aspects of the diagnostic process, such as the critical importance of the doctor-patient relationship of observer error and of formal history-taking, are best taught in the wards or outpatient clinics.

C H A P T E R 7 : STUDIES OF THYROTOXICOSIS

7.0 Summary

A non-linear discriminant function technique which, in contrast to the Bayesian model of the previous chapter, uses absolute numbers and takes some account of inter-dependence between tests is used to attempt to predict within three months the three year outcome of treatment with anti-thyroid drugs.

The general problem of treatment of thyrotoxicosis is briefly reviewed before the data, experimental design and computing technique are dealt with in some detail.

Two studies of the behaviour of patients treated with anti-thyroid drugs are described. The first detailed study of fifty patients yielded data about the time-dependent behaviour of both cured and relapsed patients. This information was used to design a second study of eighty patients, which showed that while the non-linear technique did not separate the two groups at control or three months separation, the combined values allowed 90 per cent separation with all the cure cases being identified.

The possibility of several sub-groups within the disease entity of thyrotoxicosis was also discussed in some detail.

7.1 Introduction

In/

In the previous three studies (Chapters 4-6) a sequential conditional probability model has been used where an assumption of independence between tests is made and test outcomes are divided into 2 or 3 classes (Fig. 4.1). The problems associated with the independence assumption have already been discussed (2.1.6). The division of test outcomes, e.g. age, into classes clearly does not make full use of the information provided by a test.

The model used in this chapter uses continuous data and takes some account of interdependence between tests.

7.2 The treatment of thyrotoxicosis

Although thyroid function tests have been of great value in the diagnosis of thyroid disease, their use in management is very limited. They are useful in confirming the clinical diagnosis but when it comes to deciding which of the available treatments (i.e. anti-thyroid drugs, surgery or radio-iodine therapy) is best they are of almost no value (Alexander et al, 1970).

Each treatment has a major disadvantage; thus radio-iodine has a high incidence of late hypothyroidism, while with surgery relapse and hypothyroidism are a considerable problem.

No/

No attempt will be made to review or present the relative merits of these three treatments. The decision problem is here narrowed down to that of predicting those patients who are most likely to have a remission with anti-thyroid drugs.

In the treatment of thyrotoxicosis with anti-thyroid drugs the main difficulty is that about 50 per cent of the patients relapse after completing a normal course of treatment. Since both radio-iodine therapy and surgery are destructive in type, it would be very valuable clinically to be able to predict during the early months of treatment with drugs which patients would relapse and which would remain euthyroid. If this were possible, surgery or radio-iodine could then be reserved for those who were unsuitable for drug therapy.

7.3 The studies

The patients in this study were part of a large group of thyrotoxic patients treated with anti-thyroid drugs for an average of one year and followed up for at least two more more years by Doctors W.D. Alexander and D.G. McLarty in the Gardiner Institute, University of Glasgow.

The patients were diagnosed as thyrotoxic on the basis of/

of a thyrotoxicosis index score of $\geq +19$ (Crooks et al, 1959) and a serum $PB^{127}I$ of >7.5 µG/100ml.

The patients were classified into two groups according to their response to a regime of neomercazole (30-40 mg daily, reduced to maintenance doses of 15-20 mg when euthyroid) and triiodothyronine (80 µG daily). A 'cured' group was defined as those patients who remained euthyroid (on the clinical and biochemical criteria used above) for two years after cessation of anti-thyroid treatment. A 'relapsed' group consisted of those who relapsed and became thyrotoxic (by the same criteria) at any time within two years after cessation of drug therapy.

The studies were divided into two parts:

1. First study. A group of cases was chosen on which complete data were available at control (before treatment) and at one, three, six and twelve months after the beginning of treatment (Fig. 7.1).

   This group consisted of 50 cases, 21 classified as 'cured' and 29 as 'relapsed'.

   The aims of this study were to observe the relationship of the two clusters over time and to attempt to use measurements of this relationship to discriminating between the two groups, i.e. in predicting outcome of treatment with anti-thyroid drugs.

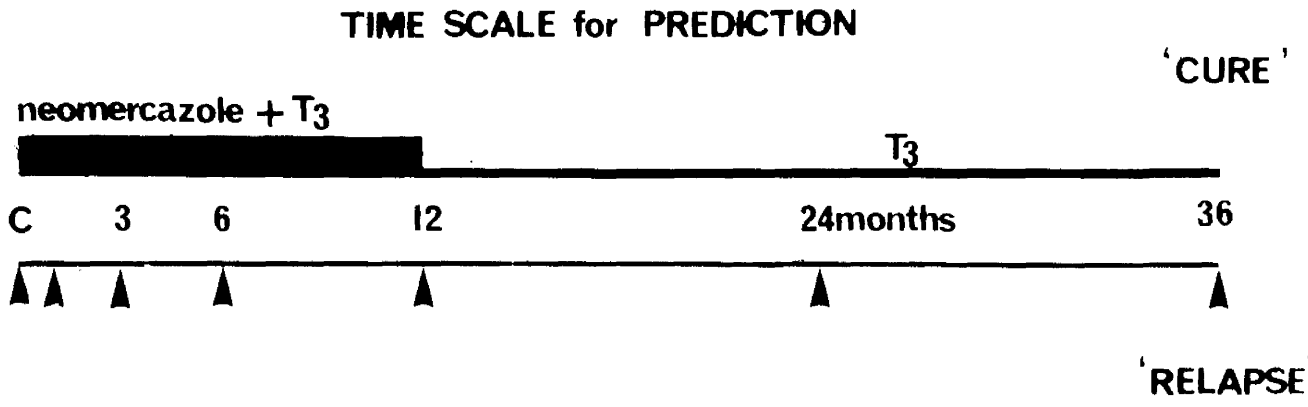


FIG. 7.1. INTERVALS USED IN STUDY.

2. Second study. A second larger group was studied after the first study was completed. It appeared likely from the results of the first study that best discrimination would be possible by combining the data from control and three months.

This group consisted of 80 cases, 40 'cured' and 40 'relapsed'. These cases were studied separately at control and at three months. This was then followed by combining control and three month data.

7.4 Data used in the studies

After much experimentation with smaller groups of tests, a total of eleven tests was used at control time and twelve tests used for all other intervals.

The list of tests used is shown in Fig. 7.2. The age was recorded in years, duration of symptoms in months and goitre size was estimated by palpation: other tests include 2 and 20 minute thyroid uptake of radioiodine ^{132}I and PBI, plasma inorganic iodide and absolute iodine uptake, T^3 resin uptake and free thyroxine index (Alexander et al, 1967, 1968).

7.5 Method of analysis

The analysis used in this study is based on the technique of non-linear discriminant function, first used by/

AGE
DURATION of symptoms
SIZE of goitre

PB ^{127}I
Plasma inorganic iodine
Absolute iodine uptake
2 min UPTAKE of ^{132}I
2–20' " " "
Thyroid CLEARANCE of ^{132}I
T^3 RESIN UPTAKE
FREE T^4 INDEX

^{132}I SUPPRESSION
(except at control)

FIG. 7.2. TESTS USED IN STUDY.

by Ferris et al (1970) and described in detail by O'Muircheartaigh (1970) in his Ph.D. thesis in Glasgow University. The programs used in the analysis are those contained in the above thesis and modified by the writer.

The technique is described in detail in Chapter 2.3 and consists, in essence, of representing the two groups of cases as two clusters in multi-dimensional space. Ellipsoids are fitted to each of these clusters and are then examined for their degree of separation. The degree of separation in this study is calculated by the likelihood ratio (2.3).

A test program is constructed from the means of the data from each group, the inverses of the co-variance matrices, the maximum values of the quadratic forms, the determinants of the two clusters and the prior probabilities for each of the two groups. All of this data is available from the analysis program except the prior probabilities, which are assumed to be equal, i.e. 0.5, 0.5.

The test program prepared by the writer will be available at a teletype in both Royal and Western Infirmaries using the KDF9 Cotan 3 on-line system. This will allow the test program to be used, if necessary, on patients attending either Thyroid Clinic.

7.6 Results of first study

The 50 cases in this part of the study were analysed using the non-linear discriminant function program already described. The data used in this analysis were the outcomes of the tests shown in Fig. 7.2. This analysis produced the five sets of likelihood ratios shown in Fig. 7.3.

When these values were plotted as scatter diagrams at each of the five intervals chosen for the study (Fig. 7.4) a variable degree of overlap between the two clusters was shown at every interval. This indicates that complete separation of the two clusters is not possible.

In assessing the information provided by the scatter diagrams, the relative stability of the clusters is worth noting. Thus, the 'cure' group remains relatively stable except at one month, when the cases are much more scattered. By twelve months the cure group is quite distinct. This may well correspond to the fact that by this time it is usually fairly obvious which cases respond well to drugs and which will not.

By contrast the 'relapse' group are very much more widely scattered at all intervals.

More detailed information can be gained by focussing attention/

"CURE"

	Control	One month	Three months	Six months	One year
1	1.173 + 4	1.956 + 7	7.382 + 5	2.294 + 8	6.042 + 29
2	2.897 + 2	1.461 + 21	2.367 + 1	1.458	3.375 + 3
3	7.978	6.342 + 4	2.197 + 1	2.570 + 1	4.581 + 5
4	6.347 + 1	1.137 + 1	7.329 + 1	1.077 + 2	4.384 + 7
5	8.813 + 3	1.558 + 2	1.741 + 2	1.027	1.784 + 2
6	3.125	3.441 - 1	4.371 + 2	1.282 + 2	5.667 + 5
7	2.901 + 2	9.585	1.114 + 1	1.845 - 1	8.629 + 3
8	Infinity	8.710 + 11	8.544 + 6	3.642	4.358 + 5
9	2.608 + 1	1.798 + 1	5.148 - 1	2.67 - 1	8.773 + 3
10	7.205 - 1	9.383	1.053 + 5	1.33 + 1	1.150 + 5
11	3.787 + 1	2.084 + 10	1.831 + 1	1.406 + 9	6.598 + 7
12	6.783	2.636 + 3	4.713 + 4	2.542	1.956 + 9
13	8.411	1.338 + 1	4.622 + 1	3.131 + 3	Infinity
14	7.298 + 1	1.035 + 1	1.364 + 3	2.906 + 1	1.456 + 6
15	2.122 + 1	1.110 + 13	6.676 + 1	1.506	1.441 + 5
16	1.546 + 4	8.448 + 6	2.549	2.999 + 2	4.211 + 5
17	7.018	9.543	2.064 + 1	1.630 + 4	2.029 + 5
18	1.175 + 4	2.009 + 17	2.372 + 1	6.409 + 5	9.474 + 4
10	6.029 + 1	1.511 + 2	1.606 + 3	3.311 - 1	6.977 + 4
20	2.409 + 5	2.170 + 8	Infinity	Infinity	2.241 + 6
21	1.702	7.237 + 10	6.911 + 1	1.994 - 1	4.211 + 4

FIG. 7.3. (a) LIST OF VALUES FOR LIKELIHOOD RATIOS IN FIRST STUDY - 'CURE' GROUP.

	"RELAPSED"				
	Control	One month	Three months	Six months	One year
1	6.350	3.800	1.196 - 4	1.556 - 7	5.804 - 24
2	1.542 - 2	1.508	1.024 - 5	8.641 - 3	6.786 - 1
3	1.655 - 2	9.739	1.029 - 4	2.229 - 3	1.791 + 5
4	2.820 - 5	2.067 + 1	9.171 - 1	1.375 - 5	1.741 + 1
5	4.064 - 6	4.181 - 5	Zero	9.480 - 9	1.571 - 2
6	3.195 - 15	4.341	2.497 - 3	4.181 - 2	1.115 + 4
7	8.858 - 22	3.328 - 13	3.208 - 9	6.373 - 14	1.967 - 9
8	4.683 - 6	7.134 - 5	2.224	4.846 - 4	zero
9	9.659 - 20	2.177	8.428 - 22	Zero	5.894 - 15
10	6.669 - 1	9.450	3.169 - 22	Zero	3.486 - 6
11	5.991 - 2	4.477 - 1	1.107 - 10	1.175 - 6	3.830 - 5
12	1.850 - 1	1.273 - 5	1.839 - 24	1.160 - 12	3.378 - 11
13	2.436 - 7	9.802 - 6	3.497 - 29	6.681 - 11	2.597 - 7
14	1.006 - 18	Zero	9.605 - 1	6.704 - 3	Zero
15	5.368 - 9	Zero	Zero	4.115 - 12	5.877 - 2
16	1.287 - 5	Zero	4.353 - 4	1.330 - 6	1.577 + 1
17	4.603 - 6	1.792 - 8	6.342 - 6	4.162 - 23	Zero
18	Zero	4.297 - 4	2.818 - 4	1.059 - 23	Zero
19	1.918 - 6	1.237 - 1	3.244 - 17	Zero	9.457 - 4
20	3.805 - 3	1.082 - 12	3.913 - 22	Zero	Zero
21	1.965	5.430 - 1	1.803 - 1	6.153 - 1	3.010 + 2
22	4.988 - 4	Zero	1.429 + 1	4.428 - 1	7.343 - 23
23	8.795 - 11	7.210 - 4	4.905 - 11	5.528 - 2	1.563 + 2
24	3.066 - 2	1.321	2.330 + 1	2.161 - 3	2.35 - 17
25	3.878 - 12	9.603 - 6	5.046 - 19	7.482 - 17	5.664 - 5
26	6.322 - 14	2.978 - 13	1.445 - 21	3.726 - 27	Zero
27	1.125 - 4	2.191 - 8	Zero	Zero	Zero
28	2.611 - 20	2.733 - 2	1.046 - 9	Zero	4.055 - 3
29	4.431 - 23	2.860 - 20	1.114 - 2	2.141 - 9	5.228 - 22

FIG. 7.3. (b) LIST OF VALUES FOR LIKELIHOOD RATIOS IN FIRST STUDY - 'RELAPSE' GROUP.

attention on the degree of overlap between the two clusters. The percentage of cases in each overlap zone and the number of actual misclassifications are shown in Fig. 7.5. The extent of the overlap is greatest at 12, 1 and 3 months respectively, while the total misclassifications are highest at 1 and 12 months. However, if major importance is attached to identifying all the cures then this is best at 12 months, even though 5 cases of 'relapse' would be misdiagnosed as cures.

Clinically it is clearly much more important to identify possible cures than possible relapses. It is obviously desirable that as many future cures be allowed to continue with a treatment which will 'cure' them even at the expense of some 'relapse' patients being treated unnecessarily. Thus, the 'cost' to the relapses is outweighed by the 'benefit' to the cures. In situations where the treatment is not as safe as with anti-thyroid drugs the problem of balancing such costs in treatment decision-making is much more complex.

An important fundamental feature of the analysis was whether the degree of separation of the two groups was stable over time. If individual cases within a cluster remain stable in their degree of separation over time, this might/

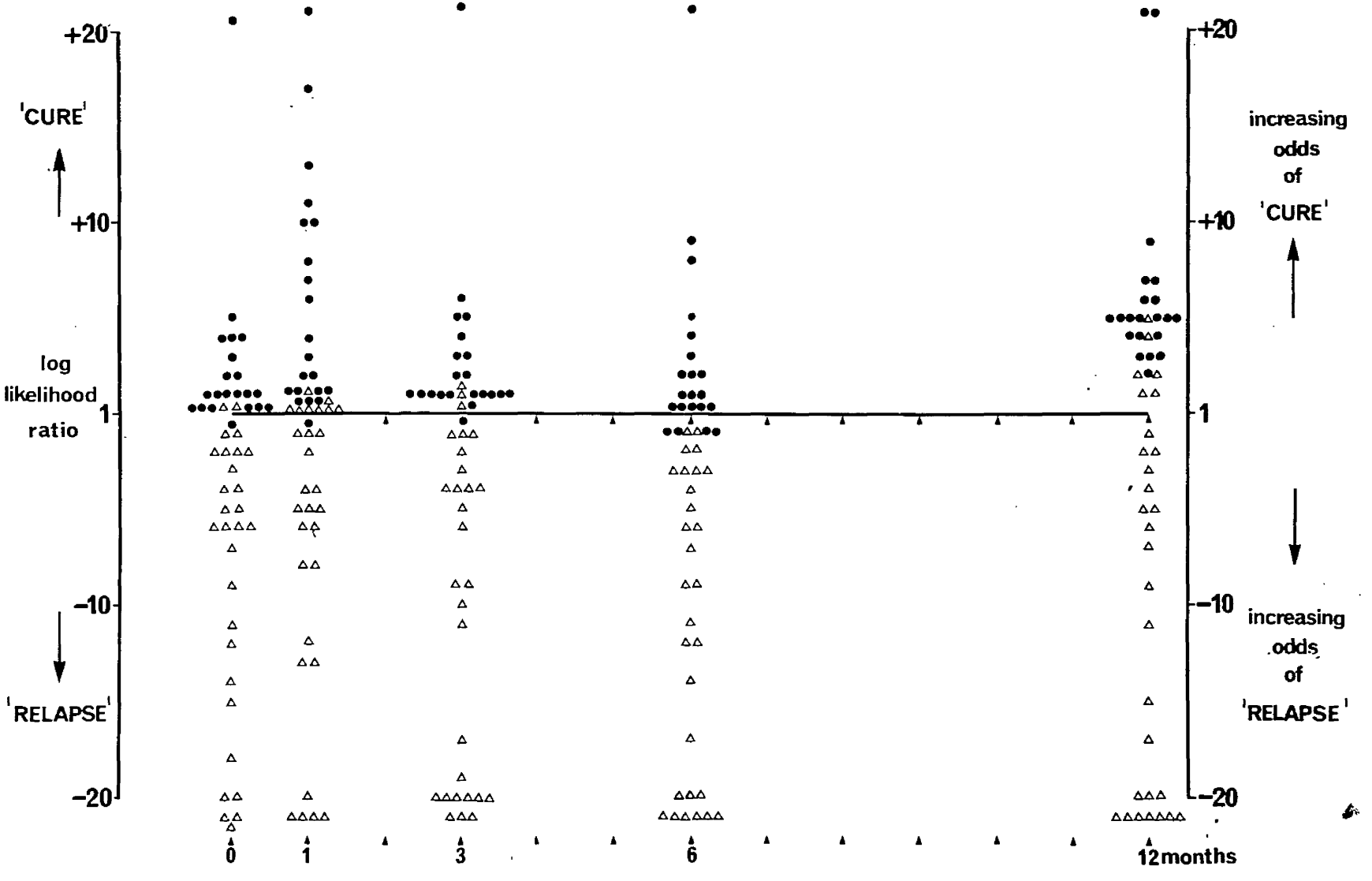


FIG. 7.4. SCATTER DIAGRAM OF LIKELIHOOD RATIOS IN FIRST STUDY.

FIRST STUDY: 21 Cures, 25 Relapses

	CONTROL	1	3	6	12 MONTHS
% in overlap	18%	32%	30%	12%	38%
MISCLASSIFICATIONS					
Cures	1	1	3	0	0
Relapsed	2	8	1	4	6
TOTAL	3	9	4	4	6
	6%	18%	8%	8%	12%

SECOND STUDY: 40 Cures, 40 Relapses

	CONTROL	3	CONTROL + THREE MONTHS
% in Overlap	66%	62.5%	22.5%
MISCLASSIFICATIONS			
Cures	8	28	8
Relapses	12	3	1
TOTAL	20	31	9
	25%	38.7%	11.25%

FIG. 7.5. MISCLASSIFICATIONS AND OVERLAP IN FIRST STUDY.

might suggest that the separation (i.e. the clusters themselves) corresponded to underlying characteristics which the cases had in common, rather than being simply a mathematical artefact. In other words, it would suggest that for example the relapse group represented a single biological entity rather than a collection of different groups whose main characteristic in common was their failure to respond to the treatment. For this reason the time dimension was added to the clusters by plotting the change in individual likelihood ratios over time.

The same scatter diagram as in Fig. 7.4 is now shown in Figs. 7.6 and 7.7 with the addition of the time vector which provides us with a dynamic picture of the relationship of the two clusters over time.

In the case of the 'cure' cluster, the cluster remains relatively stable in its relationship to the relapse cluster over time. The exception is at one month, where an almost symetrical change in likelihood ratios is apparent, compared with the control and three month values. This may well represent the 'perturbation' in the patient's thyroid system, brought about by the anti-thyroid drugs. The effect is not sustained at 3 months, perhaps because the patients are 'acclimatised' to the drug regime.

The/

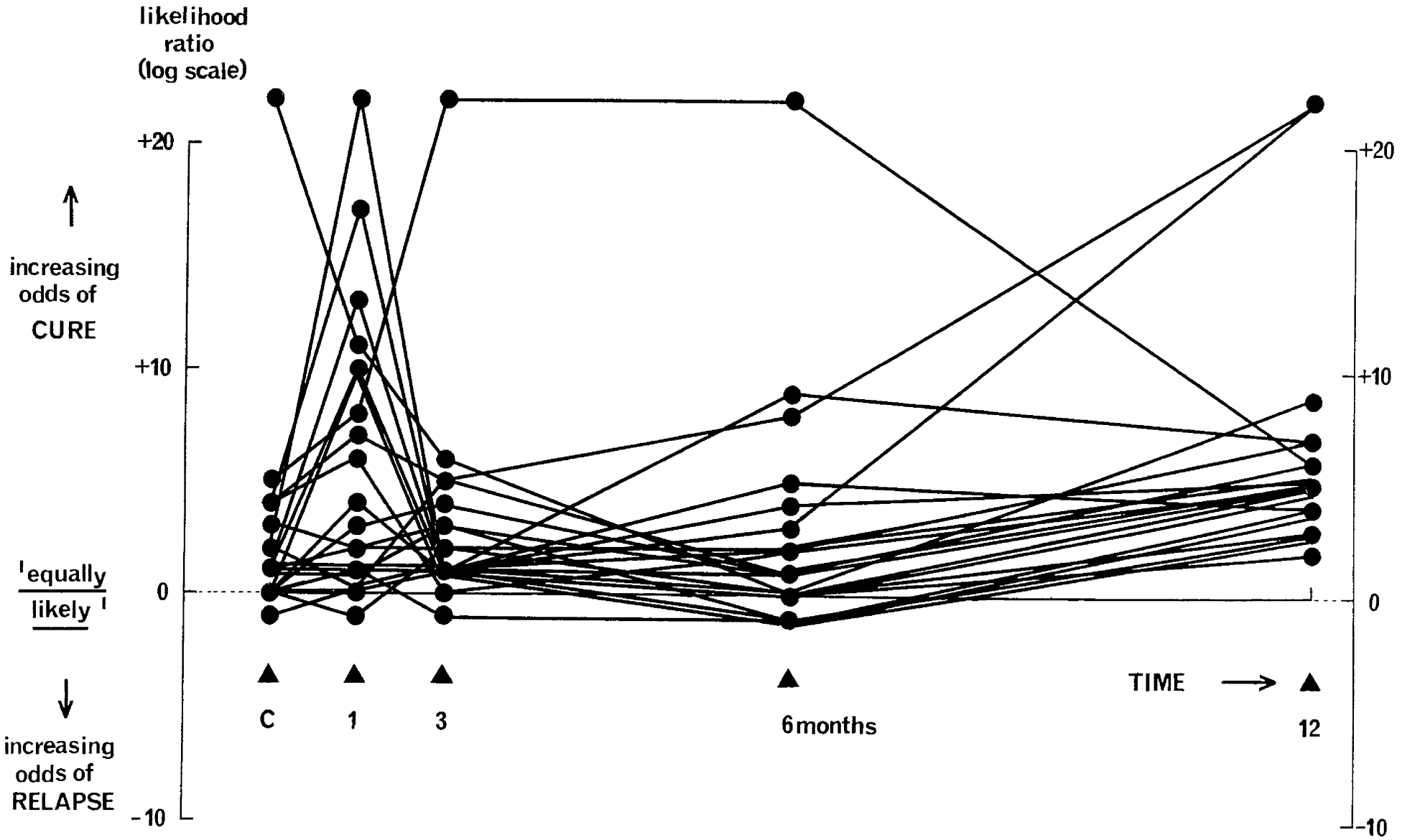


FIG. 7.6. 'DYNAMIC' GRAPH OF CHANGE IN LIKELIHOOD RATIOS OVER TIME - 'CURES'.

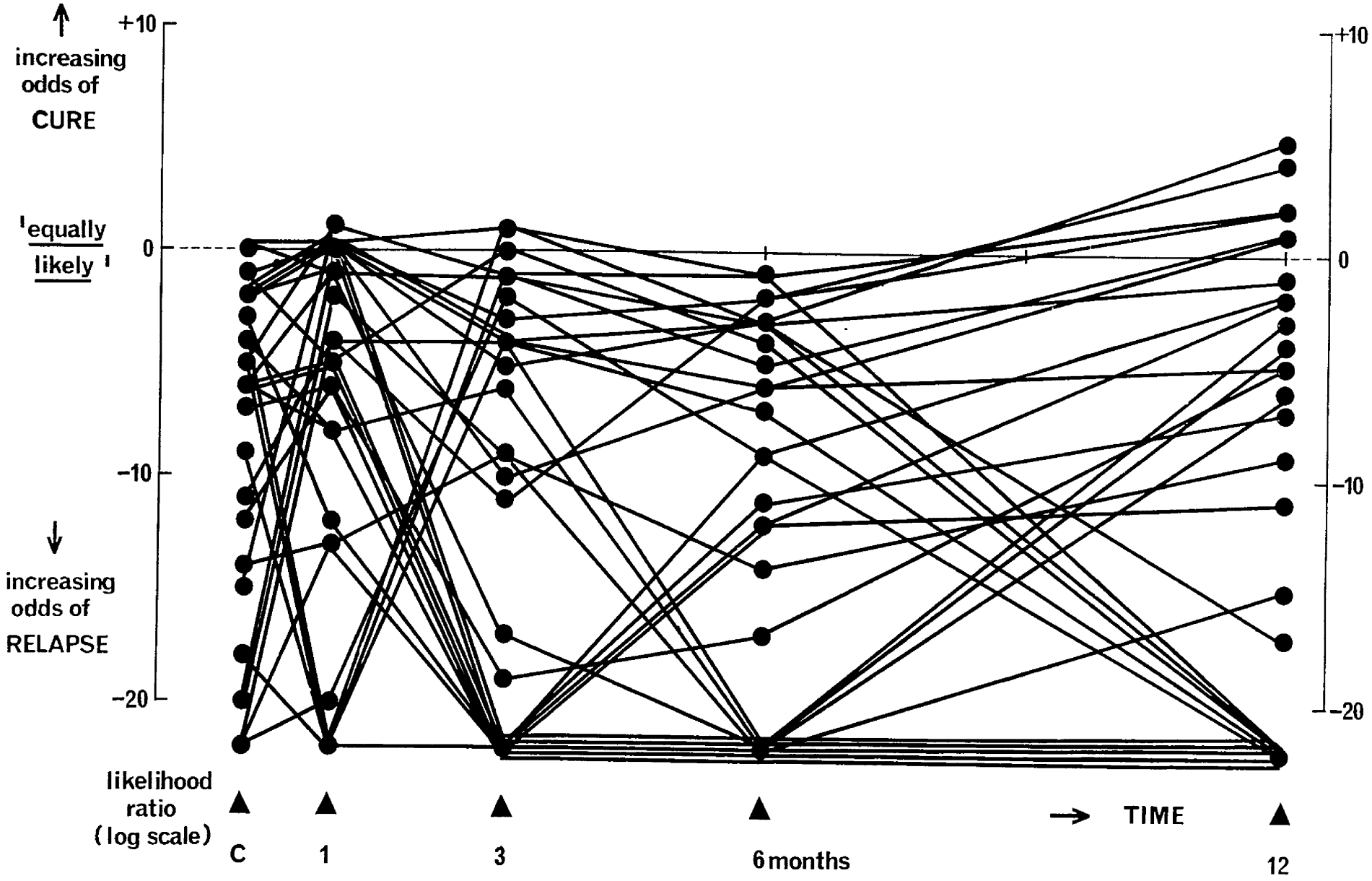


FIG. 7.7. 'DYNAMIC' GRAPH OF CHANGE IN LIKELIHOOD RATIOS OVER TIME - 'RELAPSES'.

The relative stability of the cure cluster is even more striking when it is compared to the great variability shown in Fig. 7.7 for the relapsed cluster. The exact extent of this variability can be seen by scrutinising the actual values of the likelihood ratio for any individual case, which are shown in Fig. 7.3.

When the two dynamic analyses are compared it can be seen that almost all (15/21) the cure cases remain within the 0- -10 likelihood ratio range over the 5 intervals, while in the relapse cluster the corresponding range of 0 - -10 contains only 4 out of 29 cases. In addition, only 5/29 of the relapse group had a likelihood ratio which changed >10 over the 5 intervals compared with 16/21 in the cured group. These empirical measures of cluster stability conform with the visual impression.

An interesting feature of the relapse cluster is revealed when the 5 cases (2, 3, 4, 21, 23) whose likelihood ratios are in the 0 to -10 range, are reviewed. Four of these 5 cases (2, 3, 4, 21) were those whose likelihood ratio remained in the 0 to -10 range throughout the year, while four (3, 4, 21, 23) were misclassified as cures at 12 months.

If/

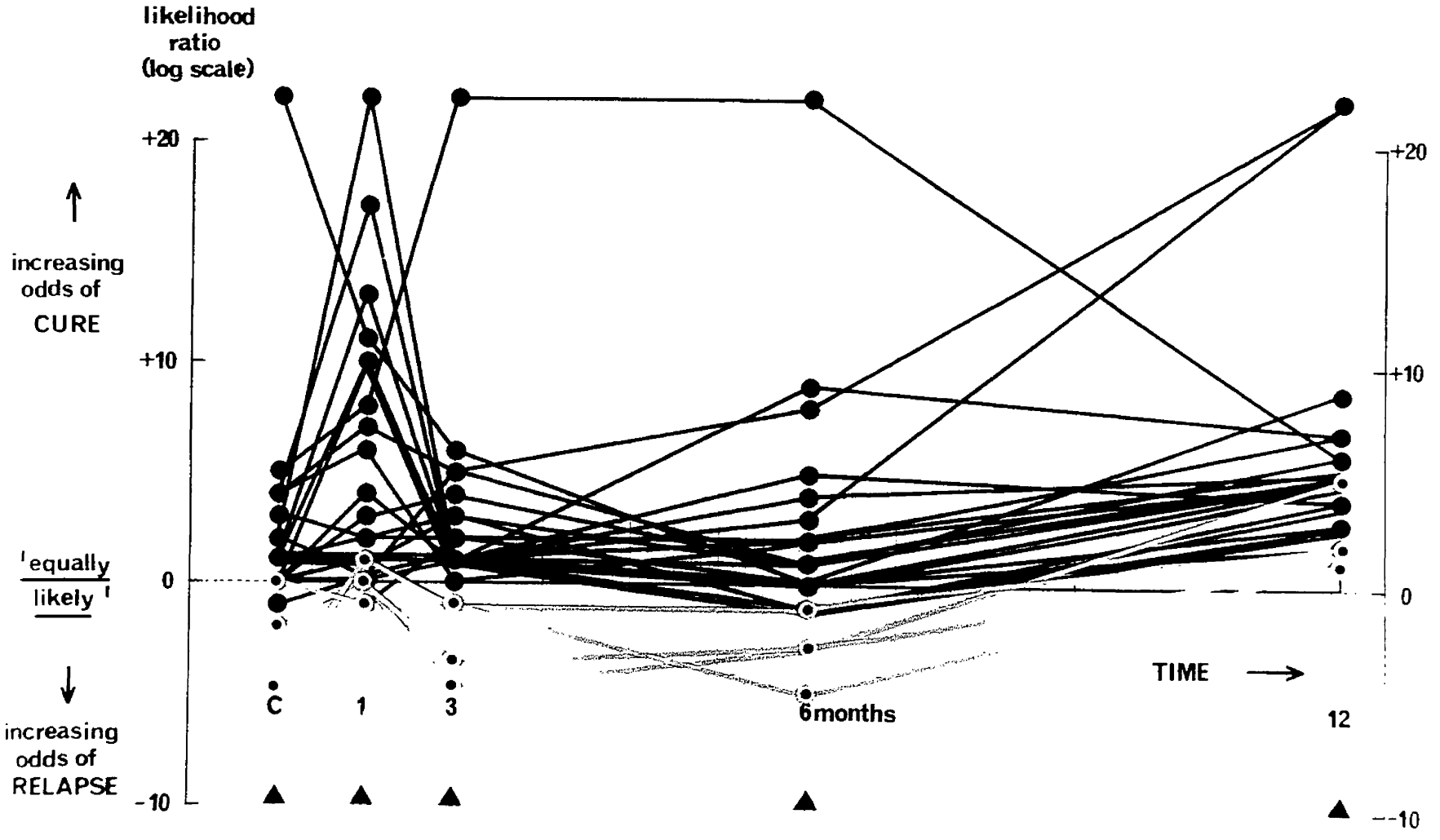


FIG. 7.8. FIVE 'STABLE' CASES IN RELAPSE GROUP.

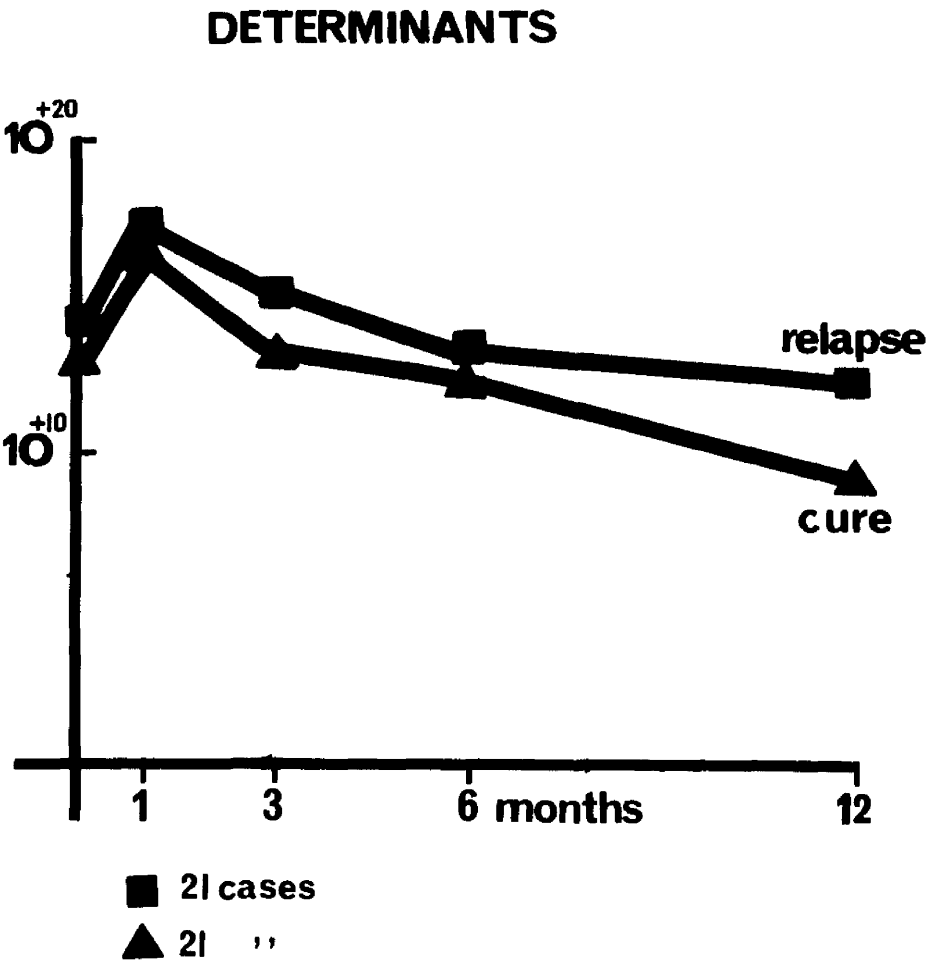


FIG. 7.9. DETERMINANTS OF CLUSTERS - A MEASURE OF DENSITY OF CLUSTERS.

If we follow this group even more closely and plot out their dynamic graph separately (Fig. 7.8) it can be seen that these five cases resemble the cure group much more closely than the relapse group. It may well be that these cases represent a group of thyrotoxics whose response to anti-thyroid drugs was dose-related and that with a higher individual dosage regime of neomercazole, they may have been cured. In other words, they might have been included in the cure cluster instead of 'hovering' around it before eventually relapsing. The implications of this interpretation of these 5 cases will be discussed later.

During the calculations of the likelihood ratios the non-linear discriminant function program also computes the determinant of each cluster. This measure can be described as the equivalent, in a multi-dimensional model, of the variance in a two dimensional model. Like the variance, it can be used as one measure of the 'spread' of a cluster.

The determinants of each cluster were plotted for the 5 intervals (Fig. 7.9). Since the number of cases in a cluster could affect its comparative 'spread', only 21 cases from each cluster were used. This analysis shows that even with the few overlying cases shown in the cure cluster (Fig.7.7)/

(7.7), this cluster is always more compact than the relapse, especially at one year. This corresponds to the spread of the likelihood ratios shown in Fig. 7.3 and may have some parallel clinically.

Thus, if a cluster is consistently more compact, it must suggest that the cases resemble one another more closely than in a 'looser' cluster and so are easier to identify as belonging to the group. If the tests used in mathematically identifying the cluster are closely enough related to the more clinical data then cases belonging to a 'tight' cluster should be easier to identify with that cluster than those in a 'looser' cluster. Clinically in this instance at one year (when this cure cluster is densenst and most distinct) it is usually easy to identify those who will relapse and those who will be cured.

7.7 Results of second study

7.7.1 Introduction

In this study it was decided to concentrate on the control and three month values in an attempt to achieve better separation and to predict outcome with new cases. The decision to use data from two intervals and to choose the control and three month values was based on the following considerations:

1./

1. Three months was considered to be the optimal time to take a decision about continuing with treatment. This is mainly because a choice between drugs and surgery is not affected by a three months course of drugs since this is commonly used before surgery to bring the thyrotoxic manifestations under control.

2. It would not be too difficult to persuade clinicians and patients to regard a three months trial of drugs as a therapeutic trial prior to a final decision being made about treatment.

3. The most striking difference between the two groups is their 'stability' over time. Since at one month both are instable, the combined control and three months values offer the possibility of emphasising the stability of the cures and the relative instability of the relapses.

4. The wide variation in individual value of tests even within groups suggested that a larger sample of each group would be desirable. This was aimed at taking as much account of this biological variability as possible.

7.7.2 Results

In the study 80 cases were used, i.e. 40 cures and 40 relapses. The analysis was confined to control and three months values separately and in combination.

The/

The scatter diagrams of the results are shown in Fig.7.10 and show considerable overlap (Fig. 7.10a,b) at the first two intervals but much better separation when they are combined (Fig. 7.10c). It must be noted that the degree of overlap is much greater with this larger sample.

7.8 Discussion

7.8.1 Introduction

The difficulties associated with the choice of treatment in thyrotoxicosis have been briefly detailed earlier (7.2) and were reviewed in detail by Hershman et al (1966). The particular problem of predicting the long-term outcome of thyrotoxicosis when treated with anti-thyroid drugs has been discussed by many authors (Solomon et al, 1953; Hershman et al, 1966; Cassidy, 1970; Alexander et al, 1970).

The approach adopted in designing this study has been to use the problem of prediction of outcome with treatment by anti-thyroid drugs as a first approach to the larger decision problem of allocating each of the three possible treatments as accurately as possible.

The data on which the study is based is derived from an/

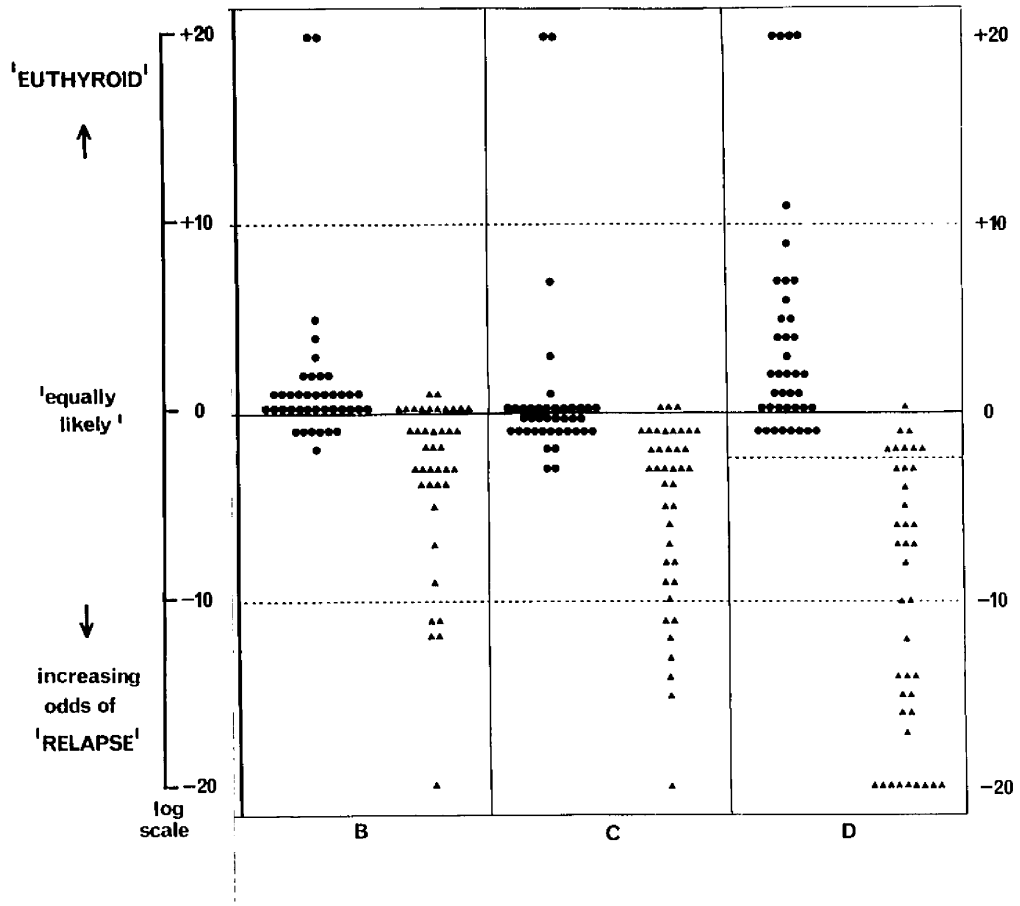


FIG. 7.10. B. SCATTER DIAGRAM FOR SECOND STUDY - CONTROL
C. SCATTER DIAGRAM FOR SECOND STUDY - THREE MONTHS
D. SCATTER DIAGRAM FOR SECOND STUDY - COMBINED WITH LINE LOWERED FOR BETTER SEPARATION OF 'CURES'.

an unselected group of thyrotoxic patients (with the rare exception of patients whose physical health made treatment a matter of urgency), all of whom were treated with the anti-thyroid drug regime already detailed (7.3).

This approach to selection would not be shared by all endocrinologists. Many might reserve drug therapy for, for example, patients who were less than 45 years of age and with a goitre of ≤60 G.

The main advantage of non-selection is that it makes the study of the natural history of the disease much easier if selection bias of this type is eliminated from the sample.

7.8.2 Choice of decision system

In all studies of computer-assisted diagnosis the choice of a suitable decision system is of vital importance to its success and its acceptance by practising clinicians. Most studies of computer-assisted diagnosis have had little impact on routine practice. An important reason for this is that the decision system studied is chosen because data is available. Other important factors are the failure to take adequate account of psychological factors or of dual pathology, while many of the problems are of little practical value. Thus, many studies in the area of thyroid disease have/

have ignored dual pathology and psychological illness and have concentrated on the calculation of the clinically trivial decision as to whether the patient is euthyroid, hyperthyroid or hypothyroid. This is a decision which can be made with a high degree of accuracy on inspection alone or, at the most, after a routine clinical assessment of the patient. Hence, to provide a novel and intellectually complex method of making such a simple clinical decision has been viewed by many clinicians as an interesting but irrelevant arithmetic exercise.

In the choice of decision problem it is therefore important to attempt to provide answers where traditional forms of decision-making are acknowledged as unsatisfactory. In other words, the aim should be to fulfil a clinical need. In doing so, it should be possible to use the same techniques to look at more fundamental problems about the nature of the diseases involved. The two approaches are indeed complementary.

The aim of this study is to provide the clinician with a method of selecting cases for anti-thyroid treatment as early as possible in the treatment of the disease. The treatment of all patients with thyrotoxicosis for a minimum of/

of three months with anti-thyroid drugs can then be regarded as a "therapeutic trial" prior to a final decision about therapy.

The data used in the study consists of a detailed battery of clinical and laboratory investigations of thyroid function which could be carried out at a centre specialising in thyroid diseases. It would be an important step in the assessment of this model to calculate the financial cost of this battery of tests if used routinely. This figure could then be balanced against the expense of treating those cases who would eventually relapse. The crucially important problem of assessing and balancing financial and non-financial costs and benefits has already been discussed in detail earlier (4.6).

The use of an on-line computer terminal makes the calculation of the decision by the test program in any individual case quite simple. A secretary can be trained in an afternoon to operate such a terminal. This would allow many clinics to use the decision model merely by telephoning data to such a secretary and being given the results within a minute over the 'phone.

7.8.3 Definitions

The last important consideration in the design of any decision/

decision model is the precise definition of the outcome categories into which the patients are to be allocated. In this study two categories were selected, i.e. 'cured' and relapsed. The exact definition of both terms has already been given (7.3).

In the choice of such categories the needs of the practising clinician must be borne in mind. Complex over-exact definitions should be avoided. By fixing the period of follow-up to three years in this study, we have ensured that the vast majority (>80 per cent) of relapses will have already occurred. The prediction, if successful, would give a useful guide to the identification of most relapses.

The tests had been performed at regular intervals over the three year follow-up period. The five intervals chosen were dictated in part by the need to have complete data and in part to yield enough information to study the biological behaviour of the two groups of cases at regular intervals.

Thus, by choosing a decision problem where conventional techniques of decision-making are unsatisfactory, by defining the outcome classes in a precise yet clinically meaningful/

meaningful way and by providing a simple method of using the technique (such as by using a trained secretary), we can greatly increase the chances of this computer-assisted decision system being widely used.

7.8.4 Analysis of results

In the analysis of results a number of important features emerge. There is a considerable overlap of varying degree at all intervals, but especially at one, three and twelve months (Fig. 7.5). This overlap is even more pronounced in the control and three month analysis of the second study, where numbers are larger. However, the overlap is greatly reduced by combining the control and three months data (Fig. 7.10). The actual percentage of misclassification is much smaller than that in the overlap and this feature also holds true in the second study.

The explanation of the much improved separation achieved by combining the control and three months data may lie, as has been suggested earlier, in the fact that the values of the likelihood ratios of the cure group remain relatively stable between these two intervals compared with the relapsed (Fig. 7.3). Another possible explanation is that the improvement is due to the increased amount of data being used in the combined analysis. Further experiments will/

will be necessary to decide which explanation is the most plausible.

Clearly, in selecting an optimal interval for discriminating between the two groups, both the degree of overlap and the number of misclassifications must be taken into account. If the overlap is not too great (as with the combined data) then it is possible to raise or lower (Fig. 7.10) the dividing line between the groups by a minor program modification. This would prevent any cures being missed, while allowing some relapses to be mistaken for cures.

The most striking feature of the scatter diagram (Fig.7.4) is that the 'cure' group (with one or two exceptions) remains stable over the five intervals. By contrast, the 'relapsed' group is much more scattered. This visual impression is confirmed by the determinants, a formal measure of the 'spread' of the clusters (Fig. 7.9).

However, this contrast between the groups becomes very much more pronounced when individual cases are followed from interval to interval. Again the cured group stay very close together, but the 'relapse' group behave in a much less coherent fashion. There is no single clear pattern in the relapse group. The immediate impression is that this/

this group is not homogeneous and that perhaps two or more patterns of time dependent behaviour are involved. It may well be that these patterns represent sub-populations or syndromes within the relapsed group.

For example, the cases which appear at the extreme of the distribution (log likelihood ratio of ≥ -20 in Fig. 7.6 and Fig. 7.7) at one interval are often much nearer the overlap at another. A few cases (17, 18, 20, 26, 27) move to the extreme at six months and stay there. These may represent a separate subgroup. Future studies will attempt to relate this group to a more severe form of the disease.

Another possible subgroup within the relapse group has already been suggested earlier (7.6.1: Fig. 7.8) which might be explained by a dose-related response to anti-thyroid drugs. Thus, if failure to respond to drugs was due in some cases to the fact that the normal dosage (7.3) was not high enough, then the percentage of cures could be increased if these cases could be identified and a more appropriate dosage regime is given. Since the distinguishing characteristic of the cure group was the relative constancy of the likelihood ratio over time, it might be possible to identify the 'potential cures' within the relapsed group by this feature/

feature. Earlier we have identified five cases whose likelihood ratio pattern resembled that of the cured group much more closely than that of the relapsed group. It is therefore of some interest that 4 out of 5 of these cases were misdiagnosed as cures at 12 months.

It must be made clear that such observations can, at this stage, be only speculative. But they demonstrate the value of this type of analysis of cluster behaviour in suggesting possible explanations of erratic and poorly understood responses to drug therapy.

It may well be that a number of syndromes exist within the disease known as thyrotoxicosis. The difficulty in predicting the behaviour of the disease and in understanding the variability in the measures of thyroid function may be because successive samples of patients may vary in their relative percentage of such syndromes.

Among the possible syndromes are:

(a) Drug-responsive cases with conventional drug regime

(b) Drug-responsive cases with other dosage regimes who are at present among the relapses

(c) Relapse patients with severe disease

(d) Other relapse patients.

The fact that the 'spread' of the relapse group, as measured/

measured by the determinants, is greater than the cure group may suggest that the latter contains one subgroup while the former contains more than one.

An important question in all these studies is whether the results of the analysis, and in particular the time-dependent behaviour, is a mathematical artefact or does in fact represent some features of the underlying biological behaviour of the disease. Once again, the evidence which we can use to support the latter hypothesis can only be circumstantial at this stage.

It would seem unlikely that the stability seen in the 'cure' cluster could be accounted for entirely by chance. In particular the almost symmetrical change in the likelihood ratios of the cure group at control, one and three months (Fig. 7.6) would be difficult to attribute to chance. It may well be due to the 'perturbation' introduced into the biological system by the metabolic 'insult' of the anti-thyroid drugs and the subsequent 'accomodation' of the system to the drugs.

At 12 months the 'cure' group is clearly separated and compact. Presumably, if a cluster is very compact this implies that cases in it resemble one another very closely. The/

The closer the resemblance, the easier it must be to recognise that any one case belongs to the group. At 12 months it is usually possible to identify those cases who will do well and those who will relapse.

7.8.5 Second study

The difference in time-dependent behaviour of the two clusters has been discussed in detail. The reasons for selecting control and three months as the best combination to use were given earlier (7.3).

The results shown in Fig. 7.10 bear out this prediction. The separation is much improved, as is the degree of overlap. It is a simple matter to lower the dividing line to ensure that all cures are clearly distinguished at the expense of 8 misclassifications among the relapses. This gives a total misclassification of 8 out of 80 (90 per cent success).

EPILOGUE

The variety of clinical applications of the technique of computer-assisted decision making used in this study is striking. While the problem area chosen is that of thyroid disease, the techniques are of wide relevance in other disease systems.

Among the major factors which have prevented their widespread use are:

1. The lack of knowledge on the part of clinicians, combined with understandable suspicion of the over-simplified claims of enthusiasts of computer medicine.
2. The choice of decision problems in the past which are of little practical interest to the clinician. Similarly, to choose as a decision problem (for calculation by computer) whether a patient is euthyroid, hypothyroid or hyperthyroid, is to employ complex techniques to solve a problem which an experienced clinician can achieve by inspection or by routine examination.
3. Computing facilities, such as on-line terminals, are not widely available and seldom in areas in hospital where the decisions are being made.

As far as future studies are concerned, the main emphasis should be on problems such as that of thyrotoxicosis treatment/

treatment studied here, where conventional decision-making is inadequate.

Using the on-line sequential probabilistic model, the writer plans to investigate the place of a computer-based screening clinic for thyroid disease in a hospital Out-patient Department. The patient would have a history taken at computer terminal, followed by a limited physical examination by carefully trained para-medical personnel. Both these items could be augmented if the computer calculates that more information from history and physical examination is needed. Several laboratory investigations are then chosen by the computer on the results of the history and examination. When these investigations are complete the patient returns to a second level clinic.

By this time the diagnosis will be clear-cut for routine cases. The difficult cases can be reviewed in detail by the clinician with much more time to spare. Patients with psychological problems (over 40 per cent of cases seen in the Thyroid Clinic at the Royal Infirmary, Glasgow) would be dealt with by a psychiatrist and a social worker who may be in attendance at this second level clinic.

The study of decision-makers will be continued with larger samples to confirm or refute the tentative conclusions of/

of the studies in this thesis. The influence of personality factors, clinical experience, 'pattern recognition' and 'channel capacity' would have to be studied in detail.

The on-line analysis and teaching system developed in Chapter 6 would have to be tested on students. It is planned to apply the same techniques to the teaching by computer of the differential diagnosis between Cushing's syndrome and obesity.

Finally, the analysis of cases of thyrotoxicosis would be continued. The two primary lines of development would be:

(a) To develop a model which took account of the time-dependent changes in the tests used in the two studies. This might make use of the differential equation approach to problems of rate changes in biological systems.

(b) To analyse further the relapse cases by the technique of numerical taxonomy to see if any clusters emerge within this group.

Needless to say (as with all investigations of clinical decision-making) techniques, models and concepts developed to solve problems in one disease system are of wide application to the whole of clinical and theoretical medicine.

A P P E N D I X A : COMPUTER PROGRAMS AND FLOW DIAGRAMS

A.1 Off-line version of the sequential probabilistic program written by the author in Algol 60.

A.2 Typical output from off-line program.

A.3 Identical off-line program written in Fortran IV.

A.4 On-line version of sequential probabilistic program written by the author in Algol 60.

FIG. A.2. TYPICAL OUTPUT FROM OFF-LINE PROGRAM.

BEST TEST	RESULT	HASHIMOTO S	SIMPLE GOITRE	C
(27) C.F. TEST	+	0.1296	0.8479	
(1) PRECIPITIN TEST	-VE	0.0394	0.9383	
(17) BE 131 I	NO DATA	0.0394	0.9383	
(28) THYMOL TURBIDITY	0.0-2.0	0.0088	0.9660	
(25) CONSISTENCY	HARD	0.0094	0.7294	
(15) RECENT INCREASE IN SIZE	NO	0.0102	0.9701	
(19) E.S.R.	21-40	0.1243	0.6853	
(21) P.B. 127 I	3.1-5.0	0.0833	0.6284	
(4) TRACHEAL DEVIATION	YES	0.0557	0.2832	
(7) CERVICAL LYMPH NODES	PALPABLE	0.0034	0.0366	
(30) AGE	60+	0.0016	0.0056	
(26) CLINICAL STATUS	EUTH.	0.0010	0.0057	
(6) FIXATION TO TISSUES	YES	0.0000	0.0010	
(5) LARYNGEAL PALSY	YES	0.0000	0.0000	
(10) HOARSENESS	YES	0.0000	0.0000	
(12) CHOKING OR TIGHTNESS	YES	0.0000	0.0000	
(16) NODULAR OR DIFFUSE	NODULAR	0.0000	0.0000	
(9) PAIN IN GOITRE	YES	0.0000	0.0000	
(13) COUGH OR STRIDOR	YES	0.0000	0.0000	
(20) 24-HOUR UPTAKE	NO DATA	0.0000	0.0000	
(23) DURATION (YEARS)	0.0-1.0	0.0000	0.0000	
(24) ESTIMATED SIZE OF GLAND	0-100	0.0000	0.0000	
(2) SERUM GLOBULINS	0.0-2.2	0.0000	0.0000	
(11) DYSPHAGIA	YES	0.0000	0.0000	
(22) P.B. 131 I AT 48 HOURS	0.21-1.0	0.0000	0.0000	
(18) GAMMAGLOBULIN	NO DATA	0.0000	0.0000	
(8) PYRAMIDAL LOBE	ABSENT	0.0000	0.0000	
(29) ZINC SULPHATE TURBIDITY	5-12	0.0000	0.0000	
(3) DISCOMFORT	YES	0.0000	0.0000	
(14) KCLO/4 DISCHARGE	NO DATA	0.0000	0.0000	

```
'BEGIN'
'REAL''ARRAY' INFJK(1:30),PRIOR,ROWSUM,Y,INF
(1:3,1:30),POST,X,LIK(1:3,1:3,1:30);
'REAL' FRED;
'INTEGER'I,J,F,Z,C,G,K,B,RESP,COUNT;
'INTEGER''ARRAY' TEST(1:30);
'PROCEDURE' SORT(INFJK);'VALUE'INFJK; 'REAL''ARRAY'INFJK;
'BEGIN'
'ARRAY' ORDER(1:2,1:30);'BOOLEAN'ST;'REAL'A;
'FOR'I=1'STEP'1'UNTIL'30'DO'
'BEGIN'
      ORDER(1,I)=INFJK(I);
      ORDER(2,I)=I
'END';
NOT DONE: ST='TRUE';
  'FOR'I=1'STEP'1'UNTIL'29'DO'
       'IF' ORDER(1,I)'GT'ORDER(1,I+1) 'THEN'
       'BEGIN'A=ORDER(1,I);
            ORDER(1,I)=ORDER(1,I+1);
            ORDER(1,I+1)=A;
            A=ORDER(2,I);
            ORDER(2,I)=ORDER(2,I+1);
            ORDER(2,I+1)=A;
            ST='FALSE'
       'END';
        'IF''NOT'ST 'THEN''GOTO' NOT DONE;
       'FOR'I=1'STEP'1'UNTIL'30'DO'
           'BEGIN'WRITE(10,G,ORDER(2,I));
                  SPACE(10,5);
                  WRITE(10,F,ORDER(1,I));
                    NEWLIN(10,1)
            'END';
'END';
OPEN(20);OPEN(10); F=LAYOUT('('ND.DDDDSSSS')');G=LAYOUT('('NDDD'
COUNT=67;
IDUMP(2,POST(1,1,1),POST(3,3,30)); EDUMP(2,INF
JK(1),INFJK(30));EDUMP(2,X(1,1,1),X(3,3,30
)); EDUMP(2,Y(1,1),Y(3,30));
EDUMP(2,PRIOR(1,1),PRIOR(3,30)); EDUMP(2,LIK(1,1,
1),LIK(3,3,30)); EDUMP(2,ROWSUN(
1,1),ROWSUM(3,30));EDUMP(2,TEST(1),TEST(30));
'FOR'I=1'STEP'1'UNTIL'3'DO'
'FOR'J=1'STEP'1'UNTIL'3'DO'
'FOR'K=1'STEP'1'UNTIL'30'DO'
  LIK(I,J,K)=READ(20);
K=READ(20); 'FOR'K=1'STEP'1'UNTIL'30'DO'
'BEGIN'TEST(K)=READ(20); PRIOR(1,K)=.566;
PRIOR(2,K)=.316;
PRIOR(3,K)=.116;'END';
 Z=30;B=30;
 'GOTO'L1;
 L4: K=READ(20);
WRITET(10,'('''('15C')'THIS'*'IS'*'NOW'
*'CASE'*'NUMBER'******')');
WRITE(10,G,K);
'FOR'K=1'STEP'1'UNTIL'30'DO''BEGIN'TEST(K)=READ(20);
PRIOR(1,K)=.566; PRIOR(2,K)=.316; PRIOR(3,K)=.116;
;'END';
COUNT=COUNT-1;
WRITET(10,'('''('15C')'RESET'*'NEW'*'
PROBABILITIES'*'AT'*'')');


WRITE(10,F,PRIOR(1,1));
WRITET(10,'('''('5C')'THIS'*'IS'*'NOW'
*'RUN '*'NUMBER'***')');
WRITE(10,G,67-COUNT);
'GOTO' L1;
 L 2;'FOR 'K=1'STEP'1'UNTIL'Z'DO'
  'BEGIN'
  INFJK(K)=0;
   PRIOR(1,K)=POST(1,RESP,C);PRIOR(2,K)=POST(2,RESP,C);
   PRIOR(3,K)=POST(3,RESP,C);
 'END';
 WRITET(10,'('''('C34S')'HASH'('6S')'SIMPLE'(
          '6S')'CANCER'('C')'')');
        WRITET(10,'('''('2C')' THE'*'NEW'*'PRIOR'*'PROBABILITIES'
         *'ARE'*'')');
        'FOR 'I=1,2,3 'DO'
        'BEGIN 'WRITE(10,F,POST(I,RESP,C));
         'END';
L1:FRED=99;B=B-1;'FOR'K=1'STEP'1'UNTIL'Z'DO'
'IF'TEST(K)'NE'0'AND'TEST(K)'NE'9'THEN'
'BEGIN'
   'BEGIN''FOR'I=1,2,3 'DO'
          'FOR'J=1,2,3 'DO'
          X(I,J,K)=PRIOR(I,K)*LIK(I,J,K);
           'FOR'J=1,2,3 'DO'
           ROWSUM(J,K)=X(1,J,K)+X(2,J,K)+X(3,J,K);
           'FOR'I=1,2,3'DO'
           'FOR'J=1,2,3'DO'
           POST(I,J,K)=X(I,J,K)/ROWSUM(J,K);
            'FOR'J=1,2,3 'DO'
'BEGIN'
               Y(J,K)=POST(1,J,K)*LN(POST(1,J,K))*.4343+
                      POST(2,J,K)*LN(POST(2,J,K))*.4343+
                      POST(3,J,K)*LN(POST(3,J,K))*0.4343;
             INF(J,K)=Y(J,K)*ROWSUM(J,K)
            'END';
            INFJK(K)=-(INF(1,K)+INF(2,K)+INF(3,K));
L3:'IF'FRED'GT'INFJK(K)'THEN'
'BEGIN'FRED=INFJK(K);F=K;
```

FIG. A-1. OFF-LINE VERSION OF THE SEQUENTIAL PROBABILISTIC PROGRAM WRITTEN BY THE AUTHOR IN ALGOL 60.

```
      CALL EXIT
      END
DATA
(1) PRECIPITIN TEST
(2) SERUM GLOBULINS
(3) DISCOMFORT
(4)TRACHEAL DEVIATION
(5) LARYNGEAL PALSY
(6) FIXATION TO TISSUES
(7) CERVICAL LYMPH NODES
(8) PYRAMIDAL LOBE
(9) PAIN IN GOITRE
(10) HOARSENESS
(11) DYSPHAGIA
(12) CHOKING OR TIGHTNESS
(13) COUGH OR STRIDOR
(14) KCLO/4 DISCHARGE
(15) RECENT INCREASE IN SIZE
(16) NODULAR OR DIFFUSE
(17) BE 131 I
(18) GAMMAGLOBULIN
(19) E.S.R.
(20) 24-HOUR UPTAKE
(21) P.B. 127 I
(22) P.B. 131 I AT 48 HOURS
(23) DURATION (YEARS)
(24) ESTIMATED SIZE OF GLAND
(25) CONSISTENCY
(26) CLINICAL STATUS
(27) C.F. TEST
(28) THYMOL TURBIDITY
(29) ZINC SULPHATE TURBIDITY
(30) AGE
+VE      -VE                0.0-2.2 2.2+              NO      YES
YES                 NO      YES                 NO    YES               IMPALP.
          ABSENT    PRESENT                   NO    YES                 NO    YES
NO      YES                 NO      YES                 NO    YES
-VE                 NO      YES                 NODULAR DIFFUSE           0-79
          0.0-0.9   0.9+                0-20    21-40   40+     0-30      31-60
0.0-3.0   3.1-5.0   5.0+      0.0-0.2   0.21-1.0  1.0+      0.0-1.0   1.1-10.0  10.0+

101-200   200+      FIRM      HARD      SOFT      HYPO.     EUTH.     HYPER.    ++
-         0.0-2.0   2.1-5.0   5.0+      5-12      13-25     25+       0-30      31-60
.7255,.3404,.9434,.7500,.9900,.9800,.9811,.8491,.9811,.8654,.8269,.7500,.9608,.6667,.3
.5577,.6923,.3514,.4583,.0909,.6316,.2308,.3478,.6226,.9057,.3208,.8372,.2553,.2222,.0
.2745,.6596,.0566,.2500,.0100,.0200,.0189,.1509,.0189,.1346,.1731,.2500,.0392,.3333,.6
.4423,.3077,.6486,.3750,.7273,.3421,.3731,.5870,.3019,.0566,.6226,.0698,.2766,.6296,.7
.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0
.0001,.0001,.0001,.1667,.1818,.0263,.0961,.0652,.0755,.0377,.0566,.0930,.4681,.1482,.2
.0010,.8148,.6666,.8889,.9900,.9020,.9600,.9608,.9348,.9167,.9375,.8125,.9556,.9334,.8
.4706,.2000,.9474,.9998,.0833,.0270,.8095,.1667,.8824,.5800,.0100,.0100,.8800,.7222,.2
.9990,.1852,.3334,.1111,.0100,.0980,.0400,.0392,.0652,.0833,.0625,.1875,.0444,.0666,.2
.5294,.8000,.0526,.0001,.7500,.4595,.1895,.5208,.0392,.0400,.8250,.0513,.1189,.2578,.6
.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0
.0001,.0001,.0001,.0001,.1667,.5135,.0010,.3125,.0784,.3800,.1650,.9387,.0001,.0200,.0
.1053,.6296,.4545,.2083,.7647,.3958,.5510,.9783,.6818,.5833,.7609,.4565,.7234,.9000,.8
.7872,.9000,.8667,.5625,.3125,.0769,.6000,.4423,.7083,.0460,.0223,.0200,.9577,.7059,.0
.8947,.3704,.5455,.7917,.2353,.6042,.4490,.0217,.3182,.4167,.2391,.5435,.2766,.1000,.1
.2128,.1000,.1333,.1563,.6250,.6923,.3500,.4231,.2500,.5300,.9677,.1081,.0323,.2841,.3
.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0
.0001,.0001,.0001,.2812,.0625,.2308,.0500,.1346,.0417,.0100,.0100,.8719,.0100,.0100,.5
,

TIME    17.38.00
```

FIG. A.3. IDENTICAL OFF-LINE PROGRAM WRITTEN IN FORTRAN IV.

```
PROGRAM ENTRY AT CHAIN 01        17.37.40        02/02/71
*XEQ
*IODUNIT20/READ
*IODUNIT70/PRINT
*STORAGE//14000
*FORTRAN
*CHAIN1

      DIMENSION NAM(4,30),IR(3,30),ICAS(30),P(30,3,3),P1(3),
     CENT(30),P2(3,3),P3(3,30),SP2(3,3),X(3),ND(30),NT(30)
C       TEST NAMES
1     FORMAT(4A8,48X)
C       TEST OUTCOMES
2     FORMAT(10A8)
C       LIKELIHOODS
3     FORMAT(15F5.4,5X)
4     FORMAT(2X,30I1,47X)
5     FORMAT(1H1,8X,9HCASE NO. ,I3,8X,9HBEST TEST,18X,
     C5HRESULT         HASHIMOTO S   SIMPLE GOITRE    CANCER)
6     FORMAT(1HO,25X,5A8,3(5X,F9.4))
C       NAMES OF TESTS
      READ(20,1)((NAM(I,J),I=1,4),J=1,30)
C       RESULTS
      READ(20,2)((IR(I,J),I=1,3),J=1,30)
C       PRIOR PROBABILITIES
      READ(20,3)(((P(I,J,K),I=1,30),J=1,3),K=1,3)
      DO 110 NC=1,67
C       CASE DATA
      READ(20,4)(ICAS(I),I=1,30)
      DO 25 I=1,18
      IF(ICAS(I)-2)15,20,20
15    ICAS(I)=ICAS(I)+1
      GO TO 25
20    ICAS(I)=9
25    CONTINUE
      DO 40 I=19,30
      IF(ICAS(I)-3)30,35,35
30    ICAS(I)=ICAS(I)+1
      GO TO 40
35    ICAS(I)=9
40    CONTINUE
      DO 45 I=1,30
45    ND(I)=0
      P1(1)=0.1
      P1(2)=0.89
      P1(3)=0.01
      DO 95 NQ=1,30
      SENT=99.0
      DO 80 I=1,30
      IF(ND(I))50,50,80
50    ENT(I)=0
      DO 65 J=1,3
      SUM=0
      Y=0
      DO 55 K=1,3
      X(K)=P1(K)*P(I,J,K)
55    SUM=SUM+X(K)
      DO 60 K=1,3
      P2(K,J)=X(K)/SUM
60    Y=Y+P2(K,J)*ALOG10(P2(K,J))
65    ENT(I)=ENT(I)-Y*SUM
      IF(SENT-ENT(I))80,80,70
70    SENT=ENT(I)

      DO 75 K=1,3
      DO 75 J=1,3
75    SP2(J,K)=P2(J,K)
      N=I
80    CONTINUE
      ND(N)=1
      NT(NQ)=N
      IF(ICAS(N)-9)85,92,92
85    J=ICAS(N)
      DO 90 I=1,3
90    P1(I)=SP2(I,J)
92    DO 94 I=1,3
94    P3(I,NQ)=P1(I)
95    CONTINUE
      PRINT 5,NC
      DO 110 N=1,30
      J=NT(N)
      K=ICAS(J)
      IF(ICAS(J)-9)100,105,105
100   IRE=IR(K,J)
      GO TO 110
105   IRE=8HNO DATA
110   PRINT 6,(NAM(I,J),I=1,4),IRE,(P3(I,N),I=1,3)
      CALL EXIT
      END
```

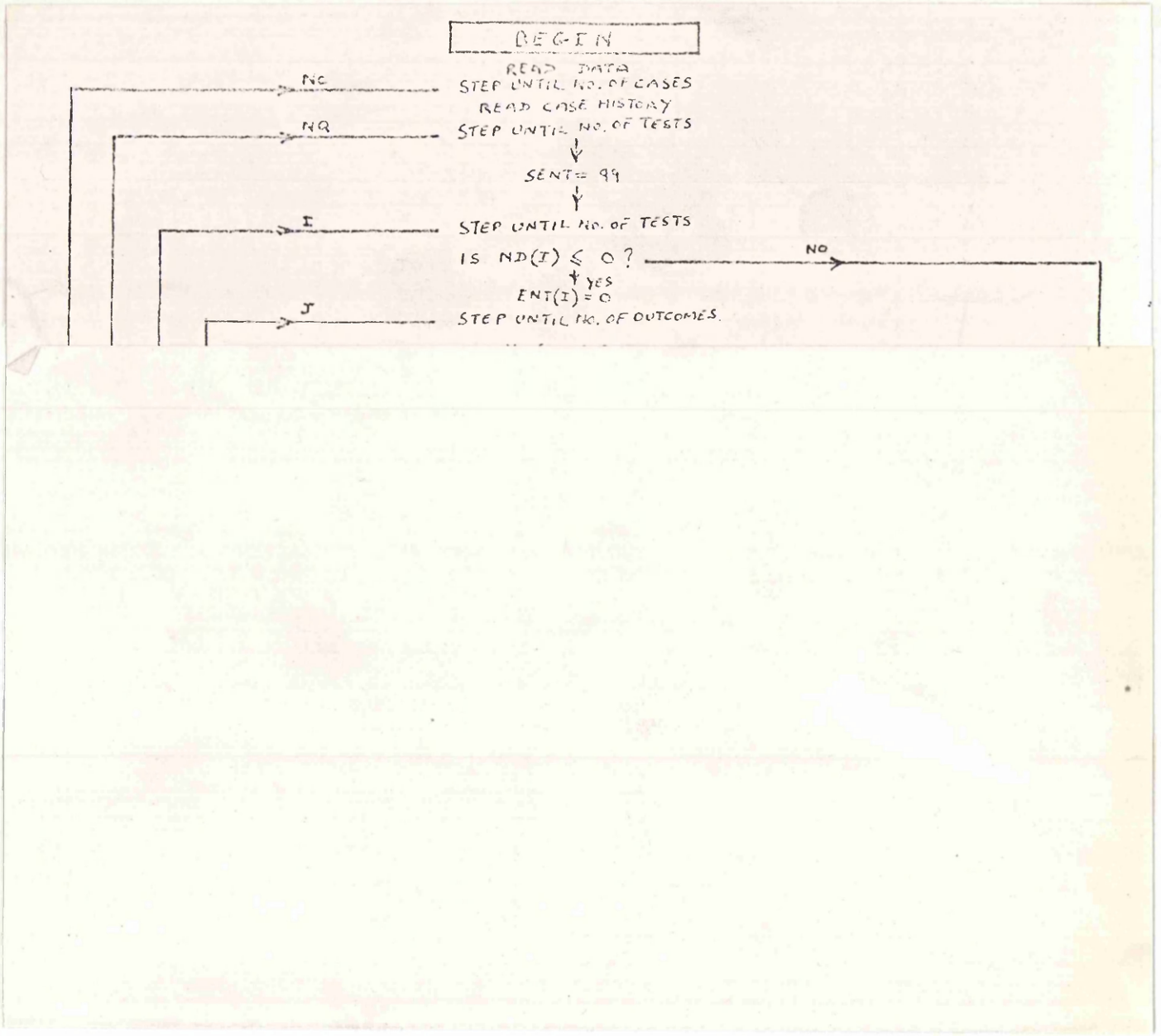
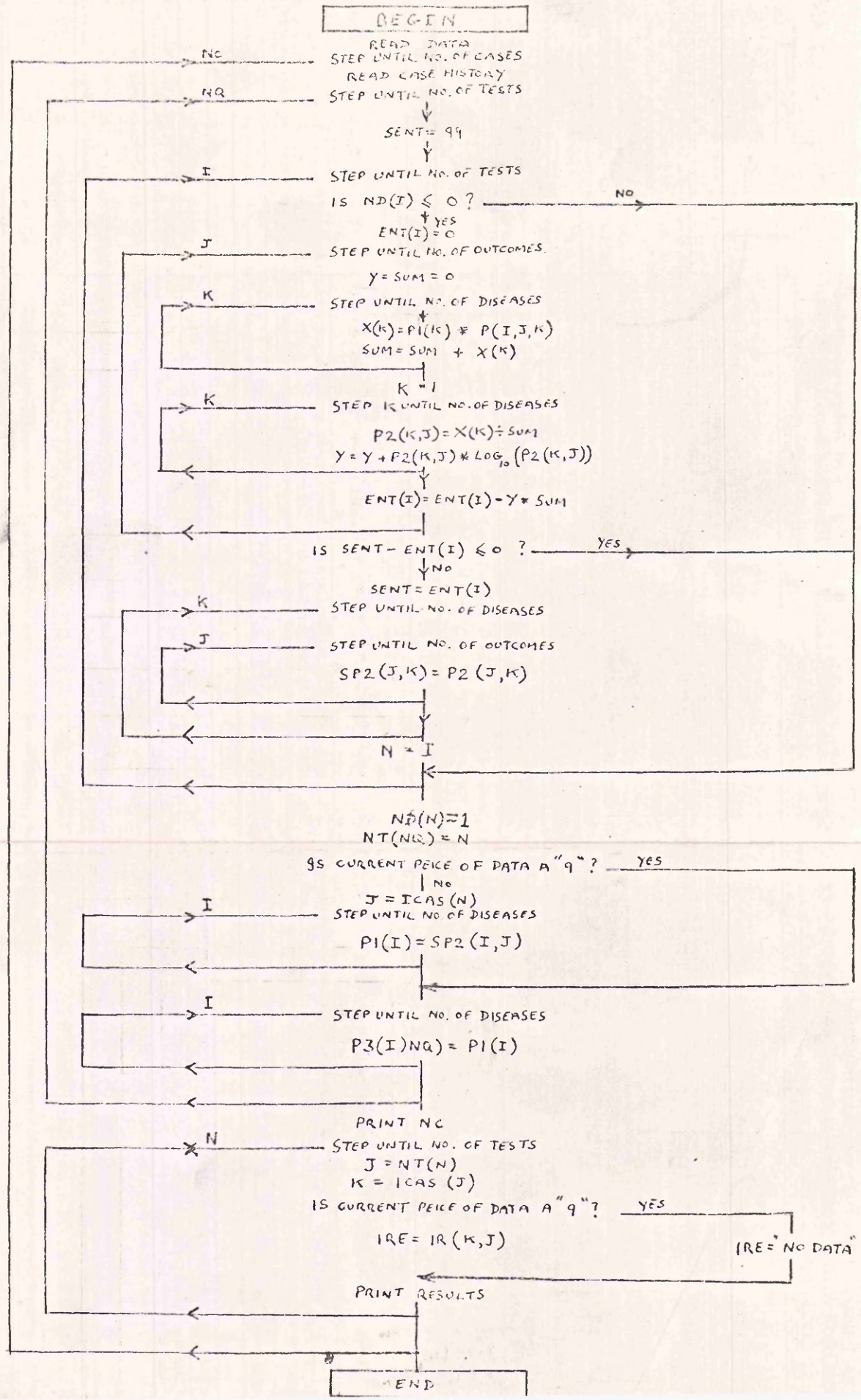


FIG. A.4. FLOW DIAGRAM FOR FORTRAN IV PROGRAM.


BEGIN
READ DATA
NC
STEP UNTIL NO. OF CASES
READ CASE HISTORY
NQ
STEP UNTIL NO. OF TESTS
SENT = 99
I
STEP UNTIL NO. OF TESTS
IS ND(I) ≤ 0 ?
NO
YES
ENT(I) = 0
J
STEP UNTIL NO. OF OUTCOMES
Y = SUM = 0
K
STEP UNTIL NO. OF DISEASES
X(K) = P1(K) * P(I,J,K)
SUM = SUM + X(K)
K = I
K
STEP K UNTIL NO. OF DISEASES
P2(K,J) = X(K) ÷ SUM
Y = Y + P2(K,J) * LOG10 (P2(K,J))
ENT(I) = ENT(I) - Y * SUM
IS SENT - ENT(I) ≤ 0 ?
YES
NO
SENT = ENT(I)
K
STEP UNTIL NO. OF DISEASES
J
STEP UNTIL NO. OF OUTCOMES
SP2(J,K) = P2(J,K)
N = I
ND(N) = 1
NT(NQ) = N
IS CURRENT PEICE OF DATA A "9" ?
YES
NO
J = ICAS(N)
I
STEP UNTIL NO. OF DISEASES
P1(I) = SP2(I,J)
I
STEP UNTIL NO. OF DISEASES
P3(I)NQ) = P1(I)
PRINT NC
N
STEP UNTIL NO. OF TESTS
J = NT(N)
K = ICAS(J)
IS CURRENT PEICE OF DATA A "9" ?
YES
IRE = IR(K,J)
IRE = " NO DATA "
PRINT RESULTS
END

```
110 ,POST[4,LIK[1:3,1:3,1:30],DAVID[1:1,1:5];
120 REAL FRED,OLDENT;
130 INTEGER I,J,N,F,Z,C,G,K,RESP,L,M;
135 INTEGER V;
136 INTEGER ANS;
140     BOOLEAN FIN,DONE;
150   FIN:=FALSE;
160   FOR K:=1 STEP 1 UNTIL 30 DO
161   BEGIN
170 PRIOR[1,K]:=PRIOR[2,K]:=PRIOR[3,K]:=0.333333
171 ; TEST[K]:=6;END;
172 FOR M:=1 STEP 1 UNTIL 5 DO
173 FOR L:=1 STEP 1 UNTIL 10 DO
174 READATA(NEW,DAVID[L,M]);
175 DATA NEW:=9,3,9,10,11,12,13,15,23,30,
176 6,6,7,8,16,25,26,0,0,0,
177 4,4,24,5,19,0,0,0,0,0,
178 6,1,2,18,27,28,29,0,0,0,
179 5,21,17,14,20,22,0,0,0,0;
181 OLDENT:=1000000.0;
190 FOR I:=1 STEP 1 UNTIL 3 DO
192 FOR J:=1 STEP 1 UNTIL 3 DO
194 FOR K:=1 STEP 1 UNTIL 30 DO
200 READATA(LIKE,LIK[I,J,K]);
205 DATA LIKE:=
210 .7255,.3444,.9434,.75,.99,.98,.9811,.8491,.9811,.8654,.8263,.75,
220 .9608,.6667,.3462,.5577,.6923,.3514,.4583,.1909,.6316,.9348,.2478,
221 .6226,
230 .9057,.3208,.3372,.0553,.2222,.0374,
240 .2745,.6596,.0566,.25,.01,.02,.3189,.1549,.0189,.1346,.1721,.25,
250 .0392,.3333,.6538,.4423,.3077,.6486,.3754,.1273,.3421,.3731,.5273,
251 .3419,
260 .0566,.6226,.0698,.9766,.6296,.7364,
270 .0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,
271 .0001,
280 .0001,.0001,.0001,.0001,.0001,.0001,.1667,.1818,.0263,.0961,.0652,
281 .0755,
290 .0377,.0566,.0930,.4681,.1482,.2274,
300 .001,.8148,.6666,.8889,.99,.902,.961,.9608,.9348,.9167,.9375,.8125,
310 .9556,.9334,.8,.4706,.2,.9474,.9998,.0833,.0270,.8095,.1667,.8824,
320 .58,.01,.01,.88,.7222,.299,
330 .999,.1852,.3334,.1111,.41,.493,.441,.4332,.3652,.0833,.1625,.1275,
340 .0444,.0666,.2,.5294,.8,.0526,.0001,.75,.4595,.1875,.5238,.8399,
350 .04,.825,.0513,.1189,.2578,.697,
360 .0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001
361 ,.0001,
370 .0001,.0001,.0001,.0001,.0001,.0001,.0001,.1667,.5135,.001,.0125,
371 .0784,
380 .38,.165,.9387,.0001,.32,.059,
390 .1053,.6296,.4545,.2083,.7647,.3958,.551,.9783,.6818,.5833,.7619,
391 .4565,.7234,
400 .9,.8936,.7872,.9,.8667,.5625,.3125,.0769,.6,.4483,.7083,.046,.7333,
410 .02,.9577,.7059,.1852,
420 .8947,.3704,.5455,.7917,.2353,.6842,.4494,.3217,.3182,.4167,.2381,
421 .5435,
430 .2766,.1,.1064,.2128,.1,.1333,.1563,.625,.6923,.33,.4231,.25,
440 .53,.9677,.1081,.0323,.2941,.3333,
450 .0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,.0001,
451 .0001,
460 .0001,.0001,.0001,.0001,.0001,.0001,.2812,.4625,.2308,.05,.1346,
461 .0417,
470 .01, .01, .8719, .01, .01, .532
1210
1220 FRED:=99;
1230 FOR L:=1 STEP 1 UNTIL 5 DO
1232 BEGIN LABEL1:DONE:=FALSE;V:=DAVID[1,L]+1;
1233 FOR M:=2 STEP 1 UNTIL V DO BEGIN
1235     K:=DAVID[M,L];
1237     IF TEST[K]=/=0 AND TEST[K]=/=9 THEN
1240         BEGIN
1240 DONE:=TRUE;
1250             FOR I:=1 STEP 1 UNTIL 3 DO
1260             FOR J:=1 STEP 1 UNTIL 3 DO
1270               X[1,J,K]:=PRIOR[I,K]*LIK[I,J,K];
1274                 FOR J:=1 STEP 1 UNTIL 3 DO ROWSUM[J,K]:=X[1,J,K]
1280                                           +X[2,J,K]+X[3,J,K];
1290                 FOR I:=1 STEP 1 UNTIL 3 DO
1400                 FOR J:=1 STEP 1 UNTIL 3 DO POST[I,J,K]:=X
1401                                           [I,J,K]/ROWSUM[J,K];
1410                 FOR J:=1 STEP 1 UNTIL 3 DO
1420               BEGIN Y[J,K]:=POST[1,J,K]*LN(POST[1,J,K])*0.4343+
1430                             POST[2,J,K]*LN(POST[2,J,K])*.4343+
1440                            POST[3,J,K]*LN(POST[3,J,K])*0.4343;
1450                INF[J,K]:=Y[J,K]*ROWSUM[J,K];
1460               END;
1470              INFJK[K]:=-(INF[1,K]+INF[2,K]+INF[3,K]);
1480                 PRINT(INFJK[K]);
1530              LABEL3: IF FRED>INFJK[K] THEN
1510                 BEGIN FRED:=INFJK[K];
1520                   C:=K;
1530                 END;
1534                   END;
1535           END; IF NOT DONE THEN GOTO LL;
1536         IF ABS(FRED-OLDENT)<0.04 THEN GOTO LL
1537         ELSE OLDENT:=FRED;
1554 FRED:=99.0;
1580   PRINT("QUESTION",C,"");
1600   READATA(TELETYPE,RESP);
1610   IF RESP=9 THEN BEGIN TEST[C]:=9;GO TO LABEL1;END;
1615   TEST[C]:=4;
1631 FOR K:=1 STEP 1 UNTIL 30 DO
1641       BEGIN PRIOR[1,K]:=POST[1,RESP,C];PRIOR[2,K]:=POST[2,RESP,C];
1651             PRIOR[3,K]:=POST[3,RESP,C];
1651       END;
1655   PRINT("THE NEW PRIOR PROBABILITIES ARE");
1661   PRINT(POST[1,RESP,C],POST[2,RESP,C],POST[3,RESP,C]);
1661 READATA(TELETYPE,ANS);
1662 IF ANS=6 THEN GOTO LL ELSE
1670 GOTO LABEL1;
1671 LL:END;
1680 END;
```

FIG. A.5. ON-LINE VERSION OF SEQUENTIAL PROBABILISTIC PROGRAM WRITTEN BY THE AUTHOR IN ALGOL 60.

ACKNOWLEDGEMENTS

I would like to acknowledge the interest and support of Professor E.M. McGirr throughout this work. Professor John Aitchison has provided generously of his time and advice in all aspects of the studies in this thesis and I am very grateful to acknowledge this. Professor W.I. Card of the Department of Medicine in Relation to Mathematics and Computing provided freely the facilities of his department and helped in the clarification of many of the ideas in the early studies in the thesis. Mr. Paul Kent prepared most of the photographs and Mrs. L. Legrix patiently typed the manuscripts.

REFERENCES

AITCHISON, J (1970)
J.Roy.Coll.Phycns.Lond. 4: 195.

AITCHISON, J. (1970)
"Choice against chance: an introduction to statistical decision theory". Addison-Wesley: Massachussets.

AITCHISON, J. (1970)
J.Roy.Statist.Soc., Series,A, 133, 206.

ALEXANDER, W.D., HARDEN, R.McG., SHIMMINS, J., McLARTY, D., McGILL, P. (1967)
Lancet, ii, 681.

ALEXANDER, W.D., HARDEN, R.McG., McLARTY, D., SHIMMINS, J. (1968) Metabolism, 18, 58.

ALEXANDER, W.D., McLARTY, D.G., ROBERTSON, J., SHIMMINS, J., BROWNLIE, B.E.W., HARDEN, R.McG. (1970)
J.Clin.Endocr. 30, 540.

ANDERSON, J.A. and BOYLE, J.A. (1968)
Brit.med.Bull. 24: 230-35.

BARKER, D.J.P., BISHOP, J.M. (1969)
Lancet, ii, 835.

BARON, D.N., FRASER, P.M. (1968)
Brit.med.Bull. 24, 236.

BILLEWICZ, W.Z., CHAPMAN, R.S., CROOKS, J., DAY, M.E., GOSSAGE, J., WAYNE, E., YOUNG, J.A. (1969)
Quart.J.Med. 38: 255.

BOYLE, J.A., GREIG, W.R., BUCHANAN, W.W., HARDEN, R.McG., FRANKLIN, D.A., McGIRR, E.M. (1966)
Quart.J.Med. 35: 565.

BOYLE, J.A. and ANDERSON, J.A. (1968)
Brit.med.Bull. 24: 224-229

BRODMAN, K., ERDMANN, A.J., LORGE, I., WOLFF, H.G. (1949)
J.A.M.A. 140: 530.

BRUCE, R.A. and COX, A.R. (1961)
North-West Med. 60:1075.

BRUCE, R.A. (1963)
in Proc.Sth.I.B.M. Medical Symp. 77.

BRUCE, R.A. and YARNALL, S.R. (1966)
J.Chronic Dis. 19: 473.

CARD, W.I. and GOOD, I.J. (1970)
Mathematical Biosciences. 6, 45.

CASSIDY, C. (1970)
Metabolism. 19, 745.

CATEL, R.B. and EBER, H.W. (1963)
Institute for Personality and Ability Testing, Illinois.

CROOKS, J., MURRAY, I.P.C., WAYNE, E.J. (1959)
Quart.J.Med. 45-28: 211.

de DOMBAL, F.T., HARTLEY, J.R., SLEEMAN (1969)
Lancet, 1, 145.

de DOMBAL, F.T., HORROCKS, J.C., STANILAND, J.R., GILL, P.W. (1971) Brit.med.J. 2: 575.

EDWARDS, D.A.W. (1970)
Scot.med.J. 15: 378.

ERDMANN, A.J. (1964)
in Jacquez, J.A.(Ed) 'The Diagnostic Process', University of Michigan Press; 215.

EYSENCK, H.J. and EYSENCK, S.B.G. (1965)
Eysenck Personality Inventory (Form B) University of London Press, London.

FEINSTEIN, A.R. (1953)
New Eng.J.Med. 269: 929.

FEINSTEIN, A.R. (1966)
Nature, 209:241.

FEINSTEIN, A.R., STERN, E.K. SPAGNUOLO, M. (1964)
Amer. Heart J. 68: 817.

FERRIS, J.B., NEVILL, A.M., BROWN, J.J., O'MUIRCHEARTAIGH, I.G. FRASER, R., ROBERTSON, J.I.S., KAY, A.W., SYMINGTON, T., LEVER, A.F. (1970)
Lancet, ii, 995.

FITZGERALD, L.T., OVERALL, J.E., WILLIAMS, C.M. (1966)
Amer.J.Roentgen. 97: 901.

FRASER, P.M. and BARON, D.N. (1968)
Proc.Roy.Soc.Med. 61, 1043.

FRIES, J.F. (1970)
Arch.Int.Med. 126: 647.

GINSBERG, A.S. and OFFENDSEN, F.L. (1968)
I.E.E.E . Trans. on Syst.Sci. and Cybernetics, SSC-4, 355.

GOOD, I.J. (1968)
Virginia J.Sci. 19, 101.

GORRY, G.A. and BARNETT, G.O. (1968)
J.A.M.A. 205: 849.

GRIBETZ, D., TALBOT, W.B., CRAWFORD, J.D. (1954)
New Engl.J.Med. 250: 555.

GROSSMAN, J., BARNETT, G.O., SWEDLOW, D. (1968)
Proc.Conf. on Eng. in Med. and Biol. 10, 471.

GROVER, N.B. and GORDON, A. (1970)
Israel J.Med.Sci. 6: 325.

GURNEY, C., OWEN, S.G., HALL, R., ROTH, M., HARPER, M., SMART, G.A. (1970)
Lancet, ii, 1275.

HAMILTON, M. (1966) 'Clinicians and Decisions',
Leeds University Press.

HARDEN, R.McG., LEVER, R., DUNN, W.R. (1969)
Lancet, i, 933.

HAYHOE, R.G.J., QUAGLINO, D., DOLL, W.R.S. (1964)
Spec.Rep.Ser.Med.Res.Coun. 304: H.M.S.O.Lond.

HEDLEY, A.J., SCOTT, A.M., WEIR, R.D. (1970)
Brit.med.J. 1: 556.

HERSHMAN, J.M. (1966)
Ann.Int.Med. 64: 1306.

HERSHMAN, J.M., GIVENS, J.R., CASSIDY, C.E., ASTWOOD, E.B.
(1966) J.clin.Endocr. 26, 803.

HOLLINGSWORTH, T.H. (1959)
Jl.R.Stat.Soc.A. 122: 221.

INGRAM, M. (1970)
Personal communication.

KLEINMUNTZ, B. (1963)
Psychol.Monogr. 77

KLEINMUNTZ, B. (1965)
Jap.Psychol.Res. 7, 189.

KLEINMUNTZ, B. (editor) (1968)
John Wiley, London. 149.

KNOX, G. (1964)
Br.J.Prev.Soc.Med. 18, 17.

KULLBACK, S. and LEIBLER, R.A. (1951)
Ann.Math.Statist. 22, 79.

LEDLEY, R.S. (1956)
J.Operations Res. Soc. Am. 4, 392.

LEDLEY, R.S. (1960)
IR.E. Med. Electronics. ME-7: 274.

LEDLEY, R.S. (1966)
J.A.M.A. 196, 933.

LEDLEY, R.S. (1969)
Proc. I.E.E.E. 57: 1900.

LEDLEY, R.S. and LUSTED, L.B. (1959)
Science. 130: 9.

LEDLEY, R.S. and LUSTED, L.B. (1960)
Operations Res. 8: 299.

LEDLEY, R.S. and LUSTED, L.B. (1962)
in Proceedings of Symposia in Applied Mathematics 14: 117.

LINDLEY, D.V. (1956)
Ann.Mathematical Statist. 27: 986.

LIPKIN, M. (1964)
Perspect.Biol.Med. 7: 485.

LIPKIN, M. and HARDY, J.D. (1958)
J.A.M.A. 166: 113.

LODWICK, G.S. (1965)
Radiol.Clin.N.Amer. 3, 487.

LUSTED, L.B. (1968)
'An Introduction to Medical Decision-making'. Charles C. Thomas, Springfield, Illinois.

MANNING, R.T. and WATSON, L. (1966)
J.A.M.A. 198, 1180.

MANTEL, H. (1967)
Cancer Research, 27, 209.

MAYNE, J.G., WEKSEL, W., SHOLTZ, P.H. (1968)
Mayo Clin.Proc. 43: 1.

MEEHL, P.E. (1946)
J.of Applied Psychology. 30: 517.

MEEHL, P.E. (1954)
Minneapolis, University of Minnesota Press.

MEEHL, P.E. (1960)
Amer.Psychologist, 15, 19.

MILLER, G.A. (1956)
Psychol.Rev. 63, 81.

NASH, F.A. (1954)
Lancet, i, 874.

NASH, F.A. (1960)
Lancet, ii, 1442.

NASH, F.A. (1963)
Proc.Roy.Soc.Med. 56, 555.

NEURATH, P.W., ENSLEIN, K., MITCHELL, G.W. (1969)
New Engl.J.Med. 280, 745.

NORRIS, R.M., BRANDT, D.W.J., LEE, A.J. (1969)
Lancet, i, 278.

NORRIS, R.M., CAUGHEY, D.E., DEEMING, L.W. (1970)
Lancet, ii, 485.

NUGENT, C.A., WARNER, H.R., DUNN, J.T., TYLER, F.H. (1964)
J.Clin.Endoc. 24: 621.

O'MUIRCHEARTAIGH, I.G. (1970)
Ph.D. Thesis, University of Glasgow.

OVERALL, J.E. and WILLIAMS, C.M. (1961)
Behav.Sci. 6: 134.

OVERALL, J.E. and WILLIAMS, C.M. (1963)
J.A.M.A. 183: 307.

OVERALL, J.E. and GORMAN, D.R. (1963)
Behav.Sci. 8, 108.

OYAMA, T. and TATSUOKA, M. (1956)
Am.Rev.Tuberc.Pulm.Dis. 73, 472.

PAYCHA, F. (1955)
Presse Med. 63: 1753.

PAYCHA, F. (1960)
Ire. Trans. on Medical Electronics. ME-7, 288.

PHILLIPS, L.D. and EDWARDS, W. (1966)
J.Exper.Psychol. 72, 346.

PRESSMAN, I. (1970)
I.E.E.E. Trans. on Educ. E-13, 134.

REALE, A., MACCACARO, G.A., ROCCA, E. (1968)
Computers and Biomedical Res. 1, 533.

RIMOLDI, H.J.A. (1961)
J.Med. Educ. 36, 73.

RINALDO, J.A., SCHEINOK, P. AND RUPE, C.E. (1963)
Ann.Internal Med. 59: 145.

ROME, H.P. (1962)
J.A.M.A. 182, 1069.

SCHEINOK, P.A. and RINALDO, J.A. (1967)
Computers, Biomed. Res. 1, 221.

SCHMIDT, S.A. (1968)
Reading, Massachussetts, Addison-Wesley.

SHANNON, C.E. (1948)
Bell.Syst.Techn.J. 27, 379.

SINES, J.O. (1970)
Brit.J.Psychiat. 116, 129.

SLACK, W.B., HICKS, G.P., REED, C.E. (1966)
New Eng.J.Med. 274, 194.

SLACK, W.V. and Van CURA L.J. (1968)
Computers and Biomed.Res. 1, 527.

SOKAL, R.R. and SNEATH, P.H.A. (1963)
'Principles of Numerical Taxonomy', W.H. Freeman, London.

SOLOMON, D.H., BECK, J.C., VANDERLAAN, W.D. (1953)
J.A.M.A. 152, 201.

STERLING, T.D., NICKSON, J., POLLOCK, S.V. (1966)
J.A.M.A. 198, 281.

SWETS, J.A. and FEURZEIG, W. (1965)
Science, 150: 572.

TAYLOR, R.T. (1967)
"The Principles of Medical Computing", Blackwell Scientific Publications, Edinburgh and Oxford.

TAYLOR, T.R. (1969)
Postgrad.Med.J. (March, 1969), 45, 173.

TAYLOR, T.R. (1970)
J.Roy.Coll.Physn. Lond. 4: 188.

TAYLOR, T.R. (1970)
Science J. 6: 81.

TAYLOR, T.R., AITCHISON, J., McGIRR, E.M. (1971)
Brit.med.J. 3, 35.

TEMPLETON, A.W., LEHR, J.L., SIMMONS, C. (1966)
Radiology, 87, 658.

TORONTO, A.F., VEASY, L.G., WARNER, H.R. (1963)
Prog.Cardiovasc.Dis. 5: 362.

WARD, J.H. (1963)
J.Amer.Stat.Ass. 58, 236.

WARNER, H.R., TORONTO, A.F., VEASEY, L.G., STEPHENSON, R. (1961) J.A.M.A. 177: 177.

WARNER, H.R., TORONTO, A.F., VEASY, L.G. (1964)
Ann. N.Y.Acad.Sci. 115: 558.

WAYNE, E.J. (1954)
Brit.med.J. 1, 411.

WILSON, W.J., TEMPLETON, A.W., TURNER, A.H., LODWICK, G.S. (1965) Radiology, 85: 1064.

WINKLER, C., HEICHERTZ, P., KLOSS, G. (1967)
Amer.J.Med.Sci. 253: 57.

WISHART, D. (1969)
Nature, 221, 97.

WORTMAN, P.M. (1966)
I.E.E.E. Trans. on Human Factors in Electronics, 8, 48.

ZIEVE, L. and HILL, E. (1955)
Gastro-enterology, 28, 785.

ZINSSER, H.H., BONNER, R.E., LEMLICH ROOTS, L. (1962)
in Proc. 4th I.B.M. Medical Symposium: I.B.M. New York, 371.

Dr. THOMAS R.
UNIVERSITY DEP
OF MEDIC
ROYAL IMFIR
86 CASTLE S
GLASGOW, C4 S

SUMMARY

MODELS OF THE DIAGNOSTIC PROCESS

The thesis describes a series of studies of decision making in clinical medicine and a mathematical framework for the investigation of the diagnostic processes advanced based on information theory, conditional probability theory and statistical decision theory. A simple description of each of these theories is presented followed by an extensive review of previous studies. The remainder of the thesis consists of three studies using a sequential probability model and a fourth study using a non-linear discriminant function model.

The sequential probabalistic model is based on a combination of Bayes' theorem and a 'minimal entropy' calculation derived from information theory. This model selects items of clinical information in such a way as to reach a diagnosis by as few steps as possible. The model had a diagnostic accuracy of 93% and used less than a third of the available tests. From these findings it is clear that there is a great deal of redundancy in the amount of information collected on patients, that tests vary widely in their discriminating power and that non-laboratory data such as the history and physical examination are in this study of comparable diagnostic value to elaborate laboratory tests such as radio-iodine studies.

The techniques used in this study promise to provide objective methods of eliminating tests of poor discriminating power and so reduce

the number of investigations performed on patients.

The sequential probability model has a diagnostic accuracy of 93% in the decision system of non-toxic goitre which was used. It is therefore as accurate as an experienced clinician. In the following chapter this model is used as a 'normative' or standard model. Six experienced clinicians tackling 20 varied cases of non-toxic goitre were compared in detail with a computer programme, using the sequential model. A diagnostic 'profile' was produced for each clinician in each case based on the discrepancies between his diagnostic performance and that of the computer programme. This profile allowed the clinicians to be compared indirectly with one another.

The most striking finding in this study was the great individual variation in diagnostic styles among the group of clinicians. Measures of personality, such as extroversion and obsessionalism also some correlation with individual elements of the clinician's diagnostic profile and may explain some of the individual differences found among the group. Obsessionalism was particularly associated with diagnostic accuracy and the number of items needed to make a diagnosis. The ability to select the most effective test at any stage in a diagnostic problem seems to be related to the number of years of general but not of specialist thyroid experience.

Individual clinicians also seem to vary in their capacity capacity to process information and this may be important in avoiding overloading clinicians with information presented to them by computer systems. Finally a technique for detecting clinicians who use a form of 'pattern recognition' was developed and it provided results suggesting that one of the six clinicians uses this skill relatively frequently.

The main importance of this study lies in the techniques which have been developed to measure some of the skills used in diagnosis. The prime value of these techniques lies in the possibility of teaching such skills; almost as important is the possibility of isolating individual skills and deciding if any of these could be satisfactorily performed by para-medical personnel. If particular types of clinical decision-making involve skills which can be taught to para-medical personnel then clinicians can be released for more complex tasks. The objective assessment of the role of para-medical personnel in a wide variety of clinical situations is likely to be of great importance in both developing countries and in western countries where the cost of traditional medical care is increasing enormously.

In the following chapter the 'normative' or standard computer diagnostic model is used to teach some of the skills involved in the diagnostic process. The 'profile' developed in the previous chapter was presented to the clinician at a computer

terminal immediately after each step in diagnosis. An interactive visual display was found to be much preferable to a standard teletype for a number of reasons. Systems of computer assisted learning were reviewed and the potential value of the 'profile' technique was discussed in relation to these.

In the final chapter a non-linear discriminant function model was used in predicting the outcome of treatment with thyrotoxicosis by antithyroid drugs. The general problem of treatment of thyrotoxicosis was briefly reviewed and two studies were described. In the first the relationship of the cluster of patients who respond to antithyroid drugs was compared with the 'relapse' cluster. Differences were noted which allowed the statistical separation of the two clusters to be achieved in the second study. This separation was possible by exploiting in the programme the fact that the drug responsive cluster was very much more stable over the first three months of treatment than the relapse cluster.

The importance of this study lies in the fact that attempts to take account of the dynamic time-relationships between groups of patients allowed successful prediction when single static measures were unsuccessful.

It is important to note that the many techniques, models and concepts developed in this thesis are of wide application to the whole of clinical and theoretical medicine.

Legends for figures

Figure	1.1	Traditional or static model of the diagnostic process.
"	1.2	Sequential decision model of the diagnostic process showing a similar recurrent cycle for history/physical examination/laboratory tests.
Figure	2.1	Example of likelihoods in non-toxic goitre showing, for example, little difference in values for pyramidal lobe among the three diseases.
"	2.2	Summary of the sequential technique.
"	2.3	Example of the calculation of an entropy value.
"	2.4a	Separation is not possible in two dimensions but is possible in three.
"	2.4b	Separation is not possible linearly but may be possible in non-linear model (lower figure).
"	2.4c	Shows the use of the non-linear cluster analysis technique. Known groups X and Y are defined by the corresponding clusters. An unknown group of five cases are then located mathematically in relation to these two clusters and the ratio of their distance from each known cluster is used to calculate the likelihood ratio. This likelihood ratio is quoted on the scale on the lower part of the diagram. The position on the scale represents the odds of the unknown cases falling into either cluster X or Y.