

Logic in Medicine

Diagnostic logic

FERGUS J MACARTNEY

The history of diagnostic logic must be as old as that of medicine itself; the motivation to improve it has probably never been stronger than it was in the mind of the caveman patient as he felt the rasp of the trepan drilling a hole in his skull. Was the hole there to let the disease out or the cure in? Doctors have always been fascinated by diagnosis and the means by which it can be reached, but until recently the purpose of studying diagnostic logic has simply been to improve thought processes. Today this remains the primary objective, but a second motive becomes ever more important—namely, that of computer modelling of the diagnostic process.

The principle of parsimony

Firstly, however, we have to consider how we may measure the usefulness of a method of diagnosis. That it should be sufficiently accurate goes almost without saying, but I would like to suggest that brevity (or, to be more specific, parsimony) is the second most important criterion by which a diagnostic process should be judged.

There is here a clear analogy with mathematical proofs. Two methods may be used to prove the same theorem, but that which always brings joy to the heart of the mathematician is the shorter and neater. This can, of course, be overdone. The great Johann Karl Friedrich Gauss, for example, when challenged as to how he had arrived at the conclusion of one of his exquisitely brief theorems, would reply loftily, "When a beautiful cathedral is built, who wants to see the scaffolding?" His proofs were indeed so short that they were often disbelieved. This did wonders for the curricula vitae of the next generation of lesser mathematicians, who published numerous lengthier proofs of what Gauss had already shown.

Is diagnosis necessary?

If we pursue the virtue of parsimony sufficiently ruthlessly we reach the interesting conclusion that under certain circumstances diagnosis is a pointless or meaningless diversion in the therapeutic process, which is, after all, the one that interests the patient. As a first example, take the question of whether a patient who has acute pain in the abdomen has appendicitis or Meckel's diverticulitis. Provided that the affected structure is in its usual position the precise diagnosis has no effect whatever on the management of the patient. It is the decision to perform a laparotomy through an incision in the right iliac fossa that matters.

As a second example, take the man aged 45 who has a blood pressure of 130/87 mm Hg. The question "Does this patient have hypertension?" is pointless, as it assumes that patients either have

or do not have hypertension, whereas all that they have is different degrees of increased blood pressure. The key questions are not diagnostic but prognostic—namely, is this patient at increased risk of death, stroke, or other complications, and if so will the benefits of antihypertensive treatment or the search for a directly treatable cause, or both, outweigh the costs?

Finally, consider the value of psychiatric diagnosis, a subject on which whole books have been written.¹ I recall as an undergraduate being alternately fascinated and astonished by the teaching of William Sargent. Into his outpatient department would come a man enshrouded by an almost palpable aura of gloom; we would be told rather unnecessarily that he was depressed. Next through the door would come a woman who looked as if she had just won a parliamentary election, even though she was anxious about the possibility of there being a recount.

"Ah, another classical example of depression," would declare the master. It took a brave student indeed to ask how two such different people could possibly be suffering from the same disease.

"Simple, the lady has atypical depression," we would be told. I have to confess that my nerve failed me at this point, and I dared ask no more. Ultimately, the truth dawned. Patients who had classical endogenous depression responded to drugs that were self evidently antidepressant. If patients who had other symptoms responded to antidepressant drugs then they must be depressed. Simple, really. However illogical the argument, Sargent was trying to maintain that what mattered was to recognise which patients would respond to different forms of treatment, rather than to argue over the truth of the diagnosis with which their problem was labelled.

Thus though I would agree with Wulff that "the clinician today must recognise that the present disease taxonomy is arbitrary, imperfect and everchanging,"² I think that it is necessary to qualify the second half of his sentence, which reads: "at the same time he must realise that we cannot do without it." This may be true in general, but it certainly is not true in particular. Furthermore, this paradox arises precisely from the arguments that Wulff, Scadding, and Campbell have put forward so well.²⁴ Disease spotting is in some respects like bird spotting, but while birds can exist in isolation diseases cannot. To be sure, a tubercle bacillus can be isolated from a patient who is sick with tuberculosis, and a ventricular septal defect is a ventricular septal defect be it inside or outside the body, but both of these are diagnosed in the first instance because of their manifestations in sick patients. We may look on our therapeutic objective as killing the tubercle bacillus or closing the ventricular septal defect, but this has no value unless we make the sick patient better. Furthermore, this disease centred approach breaks down with something like rheumatoid arthritis, as it is at present not possible to separate the disease from the patient. Our only therapeutic objective is to make the sick patient better, so this may or may not include "naming" the disease.

In a passage of great profundity, Campbell points out the following: "A disease is first recognised syndromally—a constellation of clinical features. The disease has a cause (infective, nutritional, genetic, immunological, etc); this cause produces characteristic structural changes, which in turn produce the clinical

Hospital for Sick Children, London WC1N 3JH
FERGUS J MACARTNEY, MA, FRCP, professor of paediatric cardiology

This series is edited by Professor Calbert I Phillips.

manifestations. The elucidation of the causative, structural and functional changes may not come in any particular historical order, but the paradigm has two characteristics: first, it is expected or at least hoped the relations will be specific (unique cause, unique structural and functional changes belonging to one syndrome); second, as knowledge progresses, the defining process is 'pushed to the left' in the sequence given above. In other words, a disease will not be allowed to remain in syndromal terms if it can be explained or defined in functional terms; a functional syndrome will not be left in these terms if it can be characterised structurally, and 'cause' takes priority overall."⁴

It follows that, though precise diagnosis may be unnecessary for parsimonious treatment of the patient (as I have shown), it probably remains essential for the "pushing to the left" process just described. Furthermore, though it might be imagined that precise diagnosis could be most easily dispensed with at the messy syndromal stage, which is characterised by endless futile discussion about what constitutes the syndrome (for example, depression), diagnosis can in fact be conveniently forgotten even after considerable pushing to the left, as at the functional (for example, hypertension) or even the structural stage (for example, appendicitis versus Meckel's diverticulitis).

The diagnostic process

From this point on we shall assume that diagnosis is a desirable end. Though many doctors love to imagine that diagnosis is some mystical process beyond logical analysis (this makes of them an elite priesthood), the fact is that hunch and intuition are unteachable, whereas logic can be taught. Furthermore, computers are logical but totally unimaginative.

THE BLUNDERBUSS APPROACH

The blunderbuss approach is the traditional method taught to medical students. They take a detailed history, examine the patient from top to toe, and then order every test that could conceivably have some bearing on the problem. Not until all the information is to hand do they try and work out what is the matter with the patient. This is done by fitting the pattern of abnormalities found either to textbook descriptions of diseases or to their own database of diseases in patients whom they have previously seen.

Computer applications

The nearest automated approach to the blunderbuss method is that of database comparisons.⁵ An interactive search is made of a large database of information on patients, looking for those who match the particular patient under consideration. Usually a match is first sought on a limited list of features, with the result that a rather large and inhomogeneous subset of matching patients is obtained. The number of features to be matched is then increased, and the subset usually becomes smaller. When a matching subset is obtained in which the disease diagnosed in all patients is the same the new patient is assumed to have the same disease and can then be added to the database.

This method requires the accumulation of a large database free from errors, which is very expensive, yet there is no real concept within it of a list of possible diagnoses, each having a different probability of being true. Probably the most suitable application is in the diagnosis of rare syndromes,⁶ where there is a real problem of human memory and collation of small snippets of information from diverse sources. Cases can be added from reports in journals as well as from the experience of collaborating centres, thus pooling information that could never effectively be accommodated in the memory of a single clinician.

Overall assessment

This clinical approach is woefully unimaginative, cumbersome, and extravagant (as opposed to parsimonious). When applied to laboratory tests it wastes money, not only because many of the tests originally ordered are unnecessary but also because the more tests that are ordered the more likely it is that, by chance, one or more will turn out to be "abnormal" and start a wild goose chase of further tests to investigate the chance abnormality.

If this method is so bad why does it continue to be taught? Why are students not encouraged to use searchlights rather than buckets, to borrow a phrase from Popper? Part of the reason is that a searchlight cannot be used effectively without a fairly thorough knowledge of the territory to be searched. Students need to familiarise themselves with the normal as well as the abnormal. It is good practice for them to examine the whole patient every time. A thorough history taking and physical examination is also a cheap screening test for unsuspected disease not associated with the particular problem that has brought the patient to the doctor.

A further reason for the blunderbuss method continuing to be taught is that should an unfortunate doctor ever appear before a court he is far more likely to be criticised for sins of omission than sins of commission. Until lawyers learn the virtue of parsimony and understand that medical decisions are made on the basis of uncertainty, not matters "beyond reasonable doubt," medical education will continue to be blighted accordingly.

ALGORITHMIC DIAGNOSIS

Algorithmic diagnosis will not be described in detail, as the *BMJ* has recently published series of clinical algorithms. These consist of a series of questions linked by lines labelled with the answers, which lead either to the next question, or, less often, to the diagnosis.⁷ There is one entry point to the algorithm, and if the questions are followed through a diagnosis will be reached. The same diagnosis may be reached by several different routes.

The idea of clinical algorithms comes, ironically, from a rather unfashionable method of computer programming known as flow chart construction. Flow charts were devised to mimic the Boolean logic that is "built in" to digital computers and consists in essence of manipulating the logical operators *and*, *or*, and *not* in *if* statements (for example, *if* it is snowing *and* the fountains are turned off in Trafalgar Square *then* it is New Year's Eve).

Computer applications

Algorithmic approaches to diagnosis are exemplified by programs for evaluating acid-base disorders⁸ and comatose patients.⁹ If an algorithm is sufficiently simple there is no point in computerising it; it is quicker and simpler to follow a printed version of the original algorithm.

As will be discussed, algorithms are often rather gross oversimplifications of the diagnostic process. This has led to a broadening out of the concept into production rule systems by experts in artificial intelligence.^{10,11} The comparison of databases and statistical systems requires a database of patients that is expensive and time consuming to obtain. By contrast, production rule systems require a database of knowledge consisting of production rules that are Boolean statements of the kind already described. The attraction of this approach to systems analysts is immediately obvious. Instead of spending years accumulating boring data on patients they can spend an afternoon with medical experts, picking their brains until they have translated their expertise into a series of production rules. The result has been diagnostic programs such as those for glaucoma (Casnet/glaucoma)¹² and neurological localisation in unconscious patients.¹³

Overall assessment

The great advantage of clinical algorithms is their predictability. If the same set of information is fed into them the same answer will always emerge. The problem is that real life is seldom that simple. For this reason I believe that algorithms are best devised by experts for the use of non-experts. One of the most difficult judgments that an expert has to make is how seriously to take the information and conclusions given by an unfamiliar non-expert who consults her about a patient. If the non-expert has used an algorithm with which the expert is familiar, and that algorithm is based on observations that are reproducible even in the hands of non-experts, then the expert is much better to evaluate the information. She may know, for example, that the diagnosis given by the algorithm is not certain, but at least she should be able to judge how uncertain it is.

The Boolean example of New Year's Eve given above shows one of the most serious difficulties of the production rule approach. It is not quite certain that it is New Year's Eve just because it is snowing and the fountains are turned off. Thus some measure of uncertainty is required that is propagated from one rule to another. Attempts to do this so far^{12,14} seem to be naive to statisticians,¹⁵ who after all are old hands at quantifying uncertainty. Other serious problems with this approach are the difficulties of encompassing medical knowledge in production rules¹³ and of being sure that these production rules are being applied in appropriate circumstances.¹⁴

THE HYPOTHETICODEDUCTIVE MODEL

The two models just discussed were almost diametrically opposite, but both were unsatisfactory. The key elements in the hypotheticodeductive model are the generation and testing of hypotheses, which together form a well recognised pattern of adult thinking.¹⁶ The importance of this model in diagnosis was probably first put forward by Campbell.^{17,18} It is attractive precisely because it allows for hunch and intuition in the diagnostic process, as these may be the source of hypotheses generated. None of the other models discussed here do this. Equally, this is the only model that cannot be programmed into a computer without sacrificing one of its most important characteristics. Campbell⁴ asserts emphatically that his understanding of diagnosis springs from a Popperian view of scientific discovery,¹⁹ but it is important to realise that there are some aspects in which the hypotheticodeductive model departs substantially from the Popperian approach. Although there is almost complete agreement on the importance of the generation of hypotheses and the sources of hypotheses, the Popperian view of how hypotheses should be tested is far too narrow for diagnostic purposes. These two opposing points will now be expanded.

Generation of hypotheses

The classical inductive view of science has no room for imagination but regards the foundation of scientific knowledge as consisting of particular observations (including experiments) from which general laws of the universe can be induced. The Popperian view is attractive in that it does not accept that scientists are passive observers but rather sees them as actively generating hypotheses to be tested, very often using their imaginations to do so. The generation of hypotheses is clearly identifiable in the behavioural analyses of clinicians at work^{20,21} and begins remarkably early, often as soon as the clinician is aware of the complaint, age, and sex of the patient. To give a well known example, the clinician seeing a fat fertile woman aged 40 immediately hypothesises that she has gall stones.

Once this rather romantic Popperian view of science is grasped it becomes easier to understand the subtlety of a celebrated quotation by Medawar: "In commencement addresses and other uplifting declarations, clinicians who discourse on the 'spirit of medicine' will always point out that, while there is a large and profoundly important scientific element in the practice of medicine, there is also an indefinable artistry, an imaginative insight, and medicine (they will tell us) is born of a marriage between the two. But then (it seems to me) the speaker spoils everything by getting the bride and groom confused. It is the unbiased observation, the apparatus, the ritual of fact finding and the inductive mumbojumbo that the clinician thinks of as 'scientific', and the other element, intuitive and logically unscripted, which he thinks of as a creative art."²²

Testing hypotheses

Testing hypotheses forms an important part of both the classical inductive and the Popperian deductive approaches to the foundations of scientific knowledge. While the inductive approach regards testing hypotheses as including both verification and falsification of the hypothesis, however, Popper insists that only falsification increases knowledge.¹⁹ The reason for this is that if verification is held to validate a hypothesis the assumption is that whatever experiment is being used to verify the hypothesis it will always give the same result, however often it is repeated. This is a matter of faith in the orderliness of creation, not of objectively demonstrable fact. Either you accept that it is a matter of faith or you conclude that the only acceptable way of testing a hypothesis is to prove it to be false. This is not as nihilistic as it first sounds, because scientific knowledge is advanced if an alternative hypothesis is put forward that contains more information and yet explains all the observations made so far. This new hypothesis is better because it contains more information, but it is not "the truth" because it survives only until it too is disproved.

It is clear from the observational studies already quoted that in clinical diagnosis the evaluation of hypotheses consists of both verification and falsification.²¹ Indeed, were clinicians to insist on falsifying every hypothesised diagnosis save one an enormous amount of time would be wasted, and the principle of parsimony already alluded to would be violated. Are clinicians therefore anti-Popperian? Not necessarily. It is just that when we seek to diagnose the problem in a patient we are not in the business of establishing the foundations of scientific knowledge. To do that we require certainty (that a hypothesis is false, according to Popper). From a practical point of view, to diagnose the problem we do not have to be certain that the diagnosis is correct. All we need to know is that if we manage the patient on the assumption that this diagnosis is correct the patient will do better than if

any other diagnosis is assumed. This point is foundational. Failure to understand it lies behind much muddled thinking on diagnosis.

We shall return in due course to the question of how we may be sure that a diagnosis is sufficiently accurate, but to avoid confusion it should be emphasised that what has just been written refers to the optimal management of an individual patient. Other considerations enter once we have other objectives, such as generalisations based on populations that have a particular disease. It has already been shown that diagnostic differentiation between appendicitis and Meckel's diverticulitis is not necessary for appropriate management of the patient with acute abdomen. To make generalisations about Meckel's diverticulitis, however, it is important that the diagnosis should be established as stringently as possible. Indeed, it can be argued that the less well defined a disease is the more important it is to make an accurate diagnosis.¹ The recognition by psychiatrists of the unreliability of psychiatric diagnosis led some to argue that psychiatric diagnosis should be abandoned.²³ But if this is done all hope for advancing knowledge of psychiatric illness disappears, as there is no means by which the foundation of diagnosis can be pushed to the left—that is, from the syndromic to the functional, anatomical, and causal. What is required is more reliable methods of diagnosis, not abandonment of the concept.

Sources of hypotheses

Much has been made so far of the place of imagination in generating hypotheses. In reality, most hypotheses arise from more prosaic sources, which have been well described by Cutler.²⁴ These include recognising patterns of varying degrees of complexity. There are trilogies (weight loss, exophthalmos, and tachycardia suggest thyrotoxicosis), tetralogies (squinting, hypoxic spells, cyanosis, and right ventricular hypertrophy suggest Fallot's tetralogy), and more complicated combinations of signs. These are rather amenable to computer based diagnosis, in contrast to what might be termed monologies, where the problem is diagnosed as the patient walks into the room. The facial appearances of patients who have Down's syndrome and idiopathic hypercalcaemia syndrome are so characteristic to an experienced clinician that it is doubtful whether either should be referred to as a syndrome. Certainly the term Down's syndrome should be dropped once the chromosomes have been examined, for this allows anatomical diagnosis.

Number of hypotheses

There seems to be good evidence that the limited short term memory of the clinician means that the number of hypotheses entertained at any one time is restricted to four \pm one.²⁵ Thus rejection of hypotheses is helpful not only in its own right but as a means of conserving short term memory. Once the hypothesis has been rejected it may be forgotten.

Computer modelling

The hypotheticodeductive model forms the basis of what are termed cognitive programs.²⁶ The complexity of the approach is well illustrated by Pauker *et al.*²⁷ The clearest analogue to recognising sets (for example, tetralogies) is probably in the set covering model,²⁸ which is also capable of handling the difficult problem of multiple simultaneous disorders. The most ambitious of these programs is probably Internist,²⁹ which has developed into Caduceus³⁰ and covers 500 general medical diseases and over 3500 "manifestations."

Overall assessment

Although there is much that is attractive in the hypotheticodeductive model, it has important drawbacks. Firstly, there is no real concept of the cost of evaluating hypotheses. In fact, much time and money is wasted on doing laboratory tests to rule out diagnoses that are extremely unlikely anyway or to confirm diagnoses that are already as certain as they need to be. This probably explains why the model fits history taking much more effectively than physical examination²⁶ or laboratory investigation. Questions cost nothing.

Secondly, there is no adequate explanation of the value of new information; as we shall see, this is much more subtle a matter than simply confirming or refuting hypotheses.

Thirdly, from the point of view of computer modelling (and this applies to expert systems in general) the question has to be asked whether the objective should be to mimic clinicians or to use computers to do the things that

clinicians cannot do in their heads (such as multivariate analysis). I understand the motivation of those working with computers who wish to mimic the brain, but I as a doctor want a system that will do better than the best clinician. A machine that simply does what a clinician (even a superb clinician) does is simply not a very attractive proposition.

Finally, the argument that "this is what good clinicians do; therefore it is the best model available" may not be sound. Perhaps what good clinicians do reflects how they were trained in diagnosis and they would do better if trained differently.

Bayesian probability revision

Diagnostic logic owes more than most doctors probably imagine to the work of an eighteenth century English clergyman, Thomas Bayes. When not on his pastoral duties, Bayes indulged in his hobby, mathematics; would that more clinicians were as versatile. It is impossible to discuss Bayes's theorem without introducing mathematics, but to reduce the shock I shall first give an example of diagnostic logic as it applies to my own subject, paediatric cardiology, and then show that the approach is analogous to successive applications of Bayes's theorem. This approach to diagnostic logic was learnt by myself and many others at the feet of that gifted teacher of adult cardiology, David Mendel, long before I (and probably he) knew anything about Bayes or his theorem. Similarly, I have taught this method to a generation of housemen and registrars without reference to Bayes except where some aptitude to mathematics is evident.

LEAGUE TABLE DIAGNOSIS

Phase 1—A child is referred to me as an outpatient. Into my mind comes a league table of ranked probabilities, starting with the highest. Top of the league is innocent systolic murmur, followed in order by venous hum, bicuspid aortic valve, ventricular septal defect, atrial septal defect, and patent ductus arteriosus. The ranking comes from their incidence in the population referred to me. Below patent ductus is an extremely long list, the details of which do not matter.

Phase 2—The mother brings the child in, and I observe that the child is a baby of 3 months old. This immediately knocks innocent systolic murmur, venous hum, bicuspid aortic valve, and atrial septal defect out of my top six. Ventricular septal defect and patent ductus arteriosus remain in the same order. Considerably less likely are tetralogy of Fallot and pulmonary valve stenosis, in that order, followed by everything else.

Phase 3—I introduce myself to the mother and ask her to remove all clothes from the baby except the nappy and its cover. While she is doing this I ask her a few routine questions about pregnancy and family history. These are of considerably more value in making her feel at home than making a diagnosis. To be honest, the only question worth asking from a diagnostic point of view is whether the child has had any cyanotic spells. If the answer to this were yes Fallot's tetralogy would go to the top of the list. The answer is no. When the baby is undressed he turns out to look entirely healthy and acyanotic without breathing difficulty. This hardly affects the likelihood of ventricular septal defect, as the baby is as likely to appear normal with a ventricular septal defect as without. Patent ductus arteriosus becomes slightly less likely, but pulmonary valve stenosis overtakes tetralogy because most patients at this age who have tetralogy will be cyanosed while most who have pulmonary valve stenosis will not.

Phase 4—I feel the pulses. They are normal. If coarctation had been anywhere near the top of the table it would have dropped far down. As it is, patent ductus arteriosus drops to fourth place, as even fairly small ducts are associated with jerky pulses.

Phase 5—As this is a young baby I omit observing the jugular venous pulse for the time being and palpate the heart. There is a systolic thrill maximal in the third left intercostal space. This makes patent ductus arteriosus more unlikely still but moves ventricular septal defect a little further ahead of the rest.

Phase 6—I auscultate. There are no diastolic murmurs. The systolic murmur is of ejection type. There is an expiratory ejection click at the lower left sternal border, and pulmonary closure is delayed and quiet. Pulmonary valve stenosis moves to the top of the table, and everything else moves out of sight.

Note that the diagnosis has been achieved with parsimony yet without generating or testing a single hypothesis. The limitations of short term memory are dealt with by considering in detail only the top of the league table. A cross sectional echocardiogram is now ordered with a request for Doppler interrogation of the jet through the pulmonary valve to estimate the severity of the pulmonary stenosis. If the estimated gradient is less than 30 mm Hg we will simply follow up the patient. If it is above 30 mm Hg and there is no associated atrial septal defect the child will go for cardiac

catheterisation and balloon valvuloplasty. If there is an atrial septal defect then the child will be referred for surgery without cardiac catheterisation. Thus the purpose of cross sectional echocardiography is primarily to help decision making. A secondary objective is to test the hypothesis that the patient has pulmonary valve stenosis without appreciable associated lesions.

The example given is typical of children of this age except that neither a chest x ray nor an electrocardiogram was required. If they had been performed they would simply have been regarded as phases 7 and 8, each requiring reassessment of the league table of probabilities.

BAYESIAN DIAGNOSIS

The diagnostic process will now be followed through mathematically. Let $D_1, D_2 \dots D_n$ represent all n possible mutually exclusive diagnoses present. Then $P(D_1), P(D_2) \dots P(D_n)$ are the probabilities of diagnoses D_1, D_2 , and so on. Table I gives the translations of D_1 to D_8 . These probabilities correspond to population incidences of diagnoses and total 1. Each $P(D)$ is known as a prior probability, as it is a probability assessed before obtaining any further information.

TABLE I—Translation of diagnostic notation

D_1	Innocent systolic murmur
D_2	Venous hum
D_3	Bicuspid aortic valve
D_4	Ventricular septal defect
D_5	Atrial septal defect
D_6	Patent ductus arteriosus
D_7	Tetralogy of Fallot
D_8	Pulmonary valve stenosis

At each phase of the investigation a clue to the diagnosis is given, which may be the answer to a question, the result of examining, say, the pulse, or the result of a laboratory test. The clue may be single (for example, the age of the patient) or composite (for example, the result of auscultation, which includes analysis of murmurs and sounds).

The first clue is that the patient has been referred to me, which makes it much more likely that the child has heart disease. The probability of each diagnosis now becomes a conditional probability, also known in this case as a posterior probability because it represents the probability after the clue has been given. It is represented by $P(D_1|C_1), P(D_2|C_1) \dots P(D_n|C_1)$. C_1 is clue one. The vertical line is translated as "given that." It is rather unlikely that the conditional probability will be the same as the prior probability. For example, $P(D_7)$, the population incidence of tetralogy of Fallot, is around 0.0007. $P(D_7|C_1)$, the probability that the child has tetralogy given that he has been referred to me, is 0.05. $P(D_1|C_1)$, the probability that the child has an innocent systolic murmur, given that he has been referred to me, is 0.4. These conditional probabilities are simplest obtained by looking at all outpatients referred to me over the past five years, counting the types of diagnoses, and dividing by the number of new patients seen.

The second clue is that the baby is 3 months old. The conditional probability of D_1 , innocent systolic murmur, now becomes $P(D_1|C_1 \text{ and } C_2)$, which is 0.05, because referring doctors usually adopt a "wait and see" strategy at this age in fit patients who probably have innocent systolic murmurs. On the other hand, $P(D_4|C_1 \text{ and } C_2)$ (ventricular septal defect) becomes 0.35, $P(D_5|C_1 \text{ and } C_2)$ (atrial septal defect) becomes 0.3, $P(D_7|C_1 \text{ and } C_2)$ (tetralogy) becomes 0.15, and $P(D_8|C_1 \text{ and } C_2)$ (pulmonary valve stenosis) becomes 0.08. These figures are rounded off for simplicity. Again, they are best obtained by counting previously referred outpatients, this time stratifying according to age. Note how critically these probabilities depend on the nature of my referral practice. They differ widely from the probabilities that would apply to a general practitioner or a consultant paediatrician.

From local peculiarity to recorded cases

We now want to move on, however, from the peculiarities of the local situation to the point at which we can use information gleaned from the medical reports. To simplify matters we first note that at each phase of the analysis except the first the posterior probability for the previous stage becomes the prior probability for this one. Thus we may write $P(D_1^*) = P(D_1|C_1 \text{ and } C_2)$, $P(D_2^*) = P(D_2|C_1 \text{ and } C_2)$, and so on.

Clue 3 is that the child is acyanotic and having no difficulty in breathing. What we wish to know is $P(D^*|C_3)$ for each diagnosis—for example, the

probability that the child has a ventricular septal defect given that he looks normal. Unfortunately, textbooks and original articles are not usually written in such a way as to provide this information, because it requires, in effect, a chapter entitled "Acyanotic. No respiratory distress." Such a chapter heading would not even appear in a book on differential diagnosis. To be sure, such a book would contain a chapter headed "Cyanosis," but it can almost be guaranteed that all this chapter would contain is a list of diseases, whereas we want a probability. What we may find is a chapter headed "Ventricular septal defect," and with a bit of luck we can extract from this $P(C_3|D_4^*)$, the probability that a child is acyanotic and in no distress given that he has a ventricular septal defect, is 3 months old, and has been referred to a consultant paediatric cardiologist (strictly speaking D_4^* refers to this paediatric cardiologist, but we have to assume unless there is good evidence to the contrary that the man who wrote the book is seeing the same kinds of patients as we are).

Is $P(C_3|D_4^*)$ of any use to us? Not without Thomas Bayes, for his theorem allows us to calculate $P(D_4^*|C_3)$ from $P(C_3|D_4^*)$.

For the general diagnosis D and the general clue C:

$$P(D|C) = P(C|D) \cdot P(D) / (P(C|D) \cdot P(D) + P(C|\bar{D}) \cdot P(\bar{D}))$$

\bar{D} , known as the complement of D, indicates the absence of that diagnosis. Now $P(C|\bar{D}) \cdot P(\bar{D})$ is simply the sum of the products $(P(C|D_i) \cdot P(D_i))$, where the set of D_i consists of all mutually exclusive and exhaustive diagnoses apart from the one under consideration. $P(\bar{D}) = 1 - P(D)$. Thus we may calculate $P(C|D)$ for each of five diagnoses, the four favourites and "everything else," to obtain the third column in table II. Each $P(C|\bar{D})$ corresponds to the false positive rate of a test for that diagnosis that is regarded as being positive in the absence of cyanosis or respiratory distress. It may be shown that the denominator in Bayes's expression is equal to $P(C)$, a result we shall use later.

For ventricular septal defect $P(C_3|D_4^*)$ is 0.5 and $P(C_3|\bar{D}_4^*)$ is 0.54.

$P(D_4^*|C_3) = (0.5 \times 0.35) / ((0.5 \times 0.35) + (0.54 \times 0.65))$, which gives the probability that the child has a ventricular septal defect, given that there is no cyanosis or respiratory distress. This is 0.35. The fourth column in table II show the posterior probabilities for the five diagnoses. Note that the probability of ventricular septal defect has remained unchanged, while patent ductus arteriosus has become less probable as a result of this observation. Pulmonary valve stenosis has edged ahead of tetralogy of Fallot, as expected. The alert reader will have observed that it is possible to calculate the figures in the fourth column of table II directly from those in the first and second columns without calculating the values in the third column.

At each stage of the diagnosis these probabilities are revised until pulmonary stenosis emerges as a clear favourite. Thus we see that the diagnostic process originally described bears a remarkably close resemblance to successive applications of Bayes's theorem, the differences being that ranked probabilities rather than actual probabilities are used and no calculations are consciously made. It seems likely, however, that the reordering of ranked probabilities occurs as a result of intuitive mathematics. This interpretation is strengthened when it is appreciated that Bayes's theorem can be rewritten as: posterior odds = prior odds \times likelihood ratio, where prior odds are the ratio between the probability of a diagnosis and the probability of its complement and likelihood ratio is the ratio of the probability of a positive clue in the presence and absence of the diagnosis.

Thus in the above example the prior odds on pulmonary valve stenosis were 0.08/0.92—that is, 0.087. The likelihood ratio was 0.9/0.49—that is, 1.84. The posterior odds rise to 0.087 \times 1.84—that is, 0.16. In gambling terms the odds against pulmonary valve stenosis have shortened from 11 to 1 (11 is the reciprocal of 0.087) to 6 to 1. Gambling on horses would be a great deal less popular were not punters confident of their intuitive ability to adjust gambling odds in the light of new information. One further refinement of this application is to take the natural logarithm of the likelihood ratio and call this "the weight of evidence."^{31 32}

TABLE II—Bayes's theorem applied to congenital heart disease. Note that probabilities in first and fourth columns total 1, whereas probabilities in second and third columns do not

Disease	$P(D^*)$	$P(C_3 \bar{D}^*)$	$P(C_3 D^*)$	$P(D^* C_3)$
Ventricular septal defect (D_4)	0.35	0.50	0.54	0.35
Patent ductus arteriosus (D_6)	0.30	0.40	0.58	0.23
Tetralogy of Fallot (D_7)	0.15	0.30	0.56	0.09
Pulmonary valve stenosis (D_8)	0.08	0.90	0.49	0.14
Everything else	0.12	0.95	0.47	0.22

$P(D^*)$ = Probability that diagnosis D^* is correct.

$P(C_3|\bar{D}^*)$ = Probability that child does not have cyanosis or respiratory distress given that diagnosis D^* is correct.

$P(C_3|D^*)$ = Probability that child does not have cyanosis or respiratory distress given that diagnosis D^* is incorrect.

$P(D^*|C_3)$ = Probability that diagnosis D^* is correct given that child does not have cyanosis or respiratory distress.

Bayes's theorem

A probabilistic approach is helpful not only because it models what some clinicians do but, even more importantly, it also illuminates issues that are not at all clear to many otherwise extremely competent diagnosticians. The most striking of these issues is the way in which prior probabilities can affect the usefulness of a test.

A good example is the hyperoxic test for congenital heart disease in neonates who seem to be cyanosed. This consists of letting the baby breathe 100% oxygen for 10 minutes and then measuring the systemic arterial oxygen pressure.³³ Table III shows the results. In figure 1 (a) these have been drawn as a probability tree. We take first the left hand tree. Here the prior probabilities of cyanotic heart disease, acyanotic heart disease, and lung disease have been obtained by dividing the column total of table III by the grand total. These probabilities are inserted on the top three branches of the tree. Each of these forks into two, depending on whether the $PO_2 > 150$ mm Hg or ≤ 150 mm Hg. These branches are labelled with the conditional probability calculated by dividing the numbers in each cell of table III by the column total (for example, $p(>150|cyanotic) = 2/109 = 0.02$). Notice that the probabilities on the branches of the tree are conditional on the events above them.

As has already been pointed out, the conditional probabilities on this tree are not particularly useful in diagnosis, but they are important in that we may be fairly sure that if the same test is carried out in a different environment the same conditional probabilities will emerge. In other words, the sensitivity and selectivity of the test are reasonably robust.

To obtain diagnostic information we need to invert the network, as shown in the right hand tree of figure 1 (a), to obtain probabilities conditional on the result of the test rather than on the type of disorder present. This network has two prior probabilities, corresponding to the two possible results, and three branches for each, representing the diseases. It can be obtained from table III by dividing the row totals by the grand totals to give the prior probabilities and numbers in each cell by row totals to give conditional

TABLE III—Results of hyperoxic test for congenital heart disease³³

	Congenital heart disease			Total
	Cyanotic	Acyanotic	Lung disease	
PO_2 (mm Hg):				
>150	2	153	7	162
≤ 150	107	0	16	123
Total	109	153	23	285

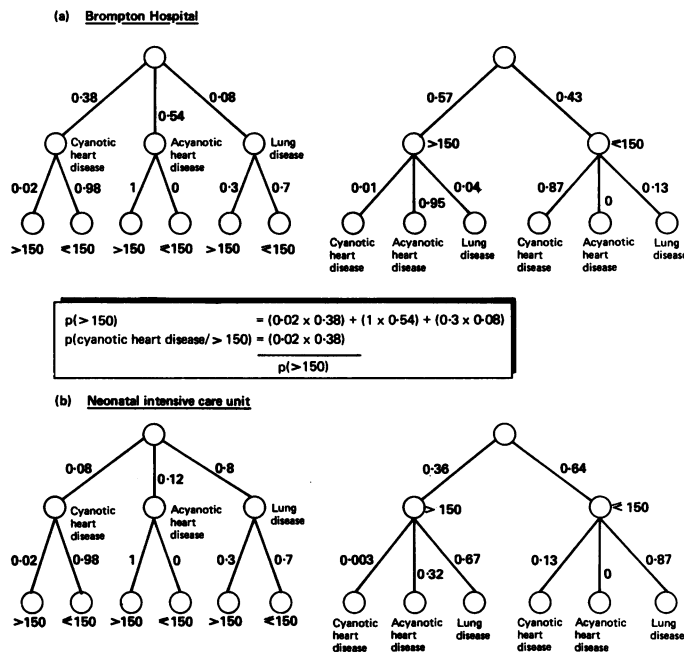


FIG 1—(a) Probability trees corresponding to results of hyperoxic test at specialist referral centre for paediatric cardiology. Equations show calculation of posterior probabilities for inversion of the probability tree. (b) Probability trees corresponding to results of hyperoxic test at neonatal intensive care unit.

TABLE IV—Posterior probabilities of diagnoses based on results of hyperoxic test

	Congenital heart disease			Lung disease
	Cyanotic	Acyanotic		
$pO_2 > 150$ mm Hg:				
Brompton	0.01	0.95		0.04
Neonatal intensive care unit	0.003	0.32		0.67
$pO_2 \leq 150$ mm Hg:				
Brompton	0.87	0		0.13
Neonatal intensive care unit	0.13	0		0.87

probabilities. Using Bayes's theorem will give the same result. It is possible to apply this directly to the network by observing that joint probabilities (for example $p(>150$ and cyanotic)) are obtained by multiplying together the probabilities on the pathways through cyanotic and >150 . To obtain $p(>150)$ we sum the joint probabilities involving >150 , which amounts to calculating the denominator in Bayes's theorem. Any conditional probability (for example, $p(\text{cyanotic} | >150)$) is then calculated by dividing the numerator in Bayes's equation (for example, $p(\text{cyanotic and } >150)$ by the denominator.

Table IV shows the posterior probabilities for the Brompton Hospital (a specialist centre for paediatric cardiology and cardiac surgery). The hyperoxic test is seen to be very useful in that if the $PO_2 > 150$ mm Hg the diagnosis is almost certainly acyanotic heart disease and highly unlikely to be cyanotic heart disease. If, on the other hand, the $PO_2 \leq 150$ mm Hg cyanotic heart disease is highly likely ($p=0.87$).

Now let us consider the performance of this test in a typical neonatal intensive care unit, where the cause of cyanosis in 80% of cases is lung disease. The left hand network of figure 1 (b) has been calculated on the assumption that cyanotic and non-cyanotic infants are distributed as for the specialist centre but make up only 20% of the whole. The probabilities conditional on the diagnosis are assumed to be the same as at the specialist centre for reasons already given. This network has then been inverted as already described to obtain the right hand network, and the probabilities conditional on the result of the test have been transferred to table IV. Observe that whatever the PO_2 the likeliest diagnosis is lung disease. Why? Because the prior probability of lung disease is so high. What is a good test in a specialist referral centre is really of no use in a neonatal intensive care unit.

The second great advantage of a probabilistic approach to diagnosis is that with little extra computational effort it can be extended into decision theory to allow the cost of an investigation to be traded off against the information it provides, thus increasing parsimony. Similarly, a decision analysis approach allows us to decide how certain we have to be about a particular diagnosis before we apply treatment on the assumption that the diagnosis is correct. As decision theory will be covered later in this series no more will be said now, but interested readers should refer to Weinstein *et al*³⁴ and Macartney *et al*.³⁵

Computer applications

The mathematical foundation of Bayes's theorem makes it an ideal computer model; probably the most successful diagnosis program yet devised for a computer was an early application of Bayes's theorem to the acute abdomen.³⁶ The computer was shown to be better than clinicians in making the diagnosis. Equally important was the fact that implementing the computerised system improved clinicians' performances by showing which clues were most helpful in discriminating diseases.³⁷ Other Bayesian systems have been described for congenital heart disease,³⁸ classification of stroke,³⁹ identifying those who will attempt suicide,⁴⁰ diagnosing solitary pulmonary nodules,⁴¹ and dyspepsia.⁴²

If the "weight of evidence" as defined above is multiplied by 100 and rounded off for convenience a rather simple means of summing the weights of evidence for and against a particular diagnosis and working back to the probability of that diagnosis can be obtained.¹⁵ The two advantages of this approach are, firstly, that it allows statistical diagnosis without a computer once the weights of evidence have been calculated and, secondly, that if a computer is used the output consists of a list of weights of evidence that are intuitively easily understood by the clinician. Thus the machine "explains" how it has come to a particular diagnosis. Approaches by artificial intelligence to diagnosis have always emphasised how important explanation is⁴³ and have criticised statistical methods for their failure to explain themselves.⁴⁴

One objection often raised to such uses of Bayes's theorem is that they assume the conditional independence of clues—that is, that the probability of a given clue in the presence of disease remains the same regardless of the presence or absence of all the other clues. Clearly this does not always apply. The probability of cyanosis given tetralogy of Fallot and finger clubbing is not the same as that given tetralogy without finger clubbing, as finger

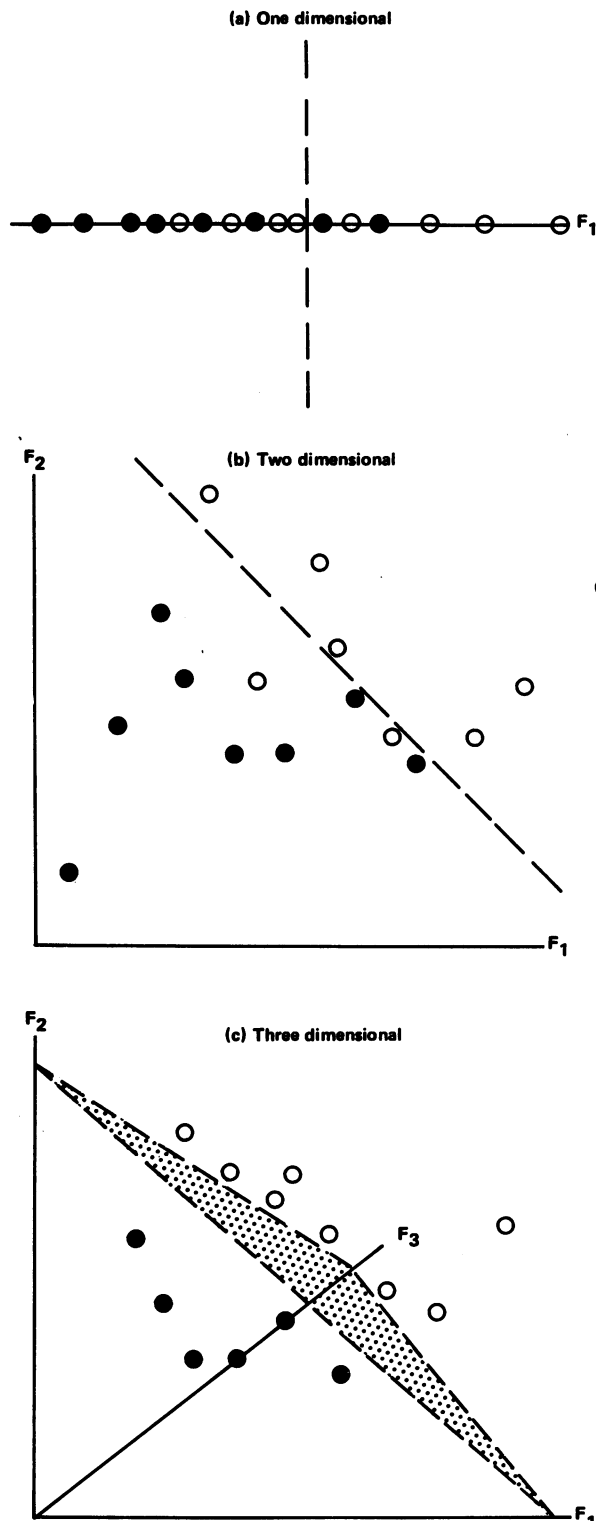


FIG 2—(a) Discriminant function analysis in one dimension. (b) Discriminant function analysis in two dimensions. (c) Discriminant function analysis in three dimensions.

clubbing and cyanosis are almost always associated beyond the age of 1 year. It is possible to apply Bayes's theorem without assuming conditional independence,⁴⁵ but the best way of dealing with the lack of conditional independence is to use multivariate analysis, as described below.

Overall assessment

As a model of clinical behaviour a Bayesian approach has been shown to be as plausible as the hypothetico-deductive, certainly in those cases where the

history is unimportant. The fact that the question of transferability—that is, whether the system developed in one environment will work in another—has been raised so often in the context of Bayesian systems^{39,46} seems to me to be a strength of the Bayesian approach rather than a weakness, as however diseases are diagnosed (unless by pathognomic clues, which are rare) prior probabilities matter, and Bayes's theorem shows why.

Alternative statistical models allowing for conditional dependence

One solution to the problem of dependence is linear discriminant function analysis, shown in figure 2. Figure 2 (a) shows the distribution of a continuous variable F_1 between two categories of patient, each with one of two diagnoses. Values of F_1 are displayed on a one dimensional line. A line at right angles intersects this line at a point whose position is calculated to maximise separation of the values of F_1 for the two sets of patients, but there is considerable overlap.

A second continuous variable F_2 is added to the model at right angles to F_1 ; this allows the results to be displayed on a two dimensional plane (fig 2 (b)). The two disease spaces are now separated by a discriminant line, again calculated to maximise separation of the two sets of patients. Overlap has been reduced but is still present.

Now F_3 is introduced at right angles to F_1 and F_2 . The observations now occupy a three dimensional space separated by a two dimensional plane (fig 2 (c)). Now complete separation of the two diagnoses has been achieved. Furthermore (as with the first two stages), posterior probabilities for a single patient can be calculated from the distance between the point corresponding to that patient and the discriminant plane.

Three dimensions is as much as the human mind can imagine, but in mathematics there is no limit to the number of dimensions and therefore no problem in conceiving a 31 dimensional space separated into two (or more) components by a 30 dimensional space. Conditional dependence is allowed for, as it alters the weights of the discriminant variables in the model.

Linear discriminant function analysis therefore provides a powerful method of computer assisted diagnosis.⁴⁷ Its use is commonly facilitated by stepwise procedures that fit first the best discriminating feature, then allow for this and fit the next best discriminator, and so on. By this means a parsimonious set of relatively few discriminators may be selected. Forward stepwise regression does not, however, necessarily pick the best set of predictors. Such a procedure does not attempt to model human diagnostic logic, although it can be explained in a way that makes sense to most clinicians.

One disadvantage of linear discriminant function analysis is that the calculation of posterior probabilities is not all that simple.⁴⁸ Multiple logistic regression has the advantage that, given the set of discriminators and their weights, the likelihood ratio is easily calculated. Another possibility is non-parametric discriminant analysis.⁴⁹

The ultimate in statistical diagnosis is to make no prior assumptions about diagnostic categories but to let the data speak for themselves. This corresponds to syndromic diagnosis by computer. For obvious reasons this approach has been mostly applied to the diagnosis of mental illness, in which there is considerable doubt about the validity of traditional diagnostic categories. On the whole, however, these multivariate methods (factor and cluster analysis) have proved to be disappointing.¹

Conclusions

No explanation of human diagnostic logic so far conceived has been entirely satisfactory, though study of the alternative models is extremely instructive. Similarly, no method of diagnosis helped by computers has been shown consistently to be superior to all others. This is an exciting field of research precisely because it is so wide open. The validation of approaches by artificial intelligence to diagnosis has been particularly scanty—either non-existent or based on fewer than 20 patients. It is essential that comparisons of alternative diagnostic aids^{44,50-52} should be carried out as stringently as are those at present for new therapeutic aids such as drugs.

FJM is supported by the Vanderbell and British Heart Foundations.

References

- Kendell RE. *The role of diagnosis in psychiatry*. Oxford: Blackwell, 1975.
- Wulff HR. *Rational diagnosis and treatment. An introduction to clinical decision-making*. 2nd ed. Oxford: Blackwell, 1981.

- Scadding JG. Diagnosis: the clinician and the computer. *Lancet* 1967;ii:877-82.
- Campbell EJM. The science of diagnosis. In: Phillips CI, Wolfe JN, eds. *Clinical practice and economics*. London: Pitman Publishing, 1977:101-12.
- Okada M, Maruyama N, Kanda T, Shirakawa K, Katagiri T. Medical data base system with an ability of automated diagnosis. *Comput Programs Biomed* 1977;7:163-70.
- Winter RM, Baraitser M, Douglas JM. A computerised data base for the diagnosis of rare dysmorphic syndromes. *J Med Genet* 1984;21:121-3.
- Williams BT. *Computer aids to clinical decisions*. Vols I and II. Florida: CRC Press, 1982.
- Bleich HL. Computer-based consultation: electrolyte and acid-base disorders. *Am J Med* 1972;53:285-91.
- Vastola EF. Assign: an automated screening system in general neurology. *Comput Biol Med* 1973;3:107-9.
- Davis R, King J. An overview of production systems. In: Elcock E, Michie D, eds. *Machine intelligence*. New York: Wiley, 1976:300-32.
- Waterman D, Hayes-Roth F. An overview of pattern-directed inference systems. In: Waterman D, Hayes-Roth F, eds. *Pattern directed inference systems*. London: Academic Press, 1978:3-22.
- Weiss S, Kulikowski C, Safir A. Glaucoma consultation by computer. *Comput Biol Med* 1978;8:25-40.
- Reggia JA. A production rule system for neurological localisation. In: *Proceedings of second annual symposium of computer applications to medical care*, New York. New York: IEEE, 1978:254-60.
- Davis R, Buchanan B, Shortliffe E. Production rules as a representation for a knowledge-based consultation program. *Artificial Intelligence* 1977;8:15-45.
- Spiegelhalter DJ, Knill-Jones RP. Statistical and knowledge-based approaches to clinical decision-support systems, with an application to gastroenterology. *Journal of Royal Statistical Society A* 1984;147(Pt 1):35-77.
- Inhelder B, Piaget J. *The growth of logical thinking from childhood to adolescence*. London: Routledge & Kegan Paul, 1958.
- Campbell EJM. Basic science, science and medical education. *Lancet* 1976;ii:134-6.
- Campbell EJM. *Clinical science. Clinical Science and Molecular Medicine* 1976;51:1-7.
- Magee B. *Popper*. London: Woburn Press, 1974.
- Elstein AS, Shulman LS, Sprafka SA. *Medical problem solving: an analysis of clinical reasoning*. Cambridge, Massachusetts: Harvard University Press, 1978.
- Kassirer JP, Gorry GA. Clinical problem solving. A behavioural analysis. *Am Intern Med* 1978;89:245-55.
- Medawar PB. *Induction and intuition in scientific thought*. London: Methuen 1969.
- Menninger K. *The vital balance: the life process in mental health and illness*. New York: Viking Press 1963.
- Cutler P. *Problem solving in clinical medicine. From data to diagnosis*. 2nd ed. Baltimore: Williams & Wilkins, 1985.
- Elstein AS, Kagan N, Shulman LS, Jason H, Loupe M. Methods and theory in the study of medical enquiry. *J Med Educ* 1972;47:85-92.
- Reggia JA, Tuhrim S. An overview of methods for computer-assisted medical decision making. In: Reggia JA, Tuhrim S, eds. *Computer-assisted medical decision making*. New York: Springer-Verlag, 1985:3-45.
- Pauker SG, Gorry GA, Kassirer JP, Schwartz WB. Towards the simulation of clinical cognition: taking a present illness by computer. *Am J Med* 1976;60:981-96.
- Reggia JA, Nau DS, Wang PY. Diagnostic expert systems based on a set covering model. *International Journal of Man-Machine Studies* 1983;19:437-60.
- Miller RA, Pople HE, Jr, Myers JD. INTERNIST-1. An experimental computer-based diagnostic consultant for general internal medicine. *N Engl J Med* 1982;307:468-76.
- Pople HE. Heuristic methods for imposing structure on ill structured problems: the structuring of medical diagnosis. In: Szolovitz P, ed. *Artificial intelligence in medicine*. Colorado: Westview Press, 1982: 119-85.
- Good IJ. Weight of evidence, corroboration, explanatory power, information and the utility of experiments. *J R Statist Soc* 1960;22(B):319-31.
- Good IJ, Card WI. The diagnostic process with special reference to errors. *Methods Inf Med* 1971;10:176-88.
- Jones RWA, Baumer JH, Joseph MC, Shinebourne EA. Arterial oxygen tension and response to oxygen breathing in differential diagnosis of congenital heart disease in infancy. *Arch Dis Child* 1976;51:667-73.
- Weinstein MC, Fineberg HV, Elstein AS, et al. *Clinical decision analysis*. Philadelphia: WB Saunders, 1980.
- Macartney FJ, Douglas J, Spiegelhalter D. To catheterise or not to catheterise? An approach based on decision theory. *Br Heart J* 1984;51:330-8.
- De Dombal FT, Leaper DJ, Staniland JR, McCann AP, Horrocks JC. Computer-aided diagnosis of acute abdominal pain. *Br Med J* 1972;ii:9-13.
- De Dombal FT, Leaper DJ, Horrocks JC, Staniland JR, McCann AP. Human and computer aided diagnosis of abdominal pain: further report with emphasis on performance of clinicians. *Br Med J* 1974;ii:376-80.
- Warner HR, Toronto AF, Veasy LG, Stephenson RS. A mathematical approach to medical diagnosis: application to congenital heart disease. *JAMA* 1961;177:177-83.
- Zagoria R, Reggia J. Transferability of medical decision support systems based on Bayesian classification. *Med Decis Making* 1983;3:501-9.
- Gustafson DH, Greist JH, Stauss FF, Erdman H, Laughren T. A probabilistic system for identifying suicide attemptors. *Comput Biomed Res* 1977;10:83-9.
- Templeton AW, Jansen C, Lehr JL, Huff R. Solitary pulmonary lesions. Computer-aided differential diagnosis and evaluation of mathematical methods. *Radiology* 1967;89:605-13.
- Horrocks JC, De Dombal FT. Computer aided diagnosis of "dyspepsia." *American Journal of Digestive Diseases* 1975;20:397-406.
- Teach RL, Shortliffe EH. An analysis of physician attitudes regarding computer-based clinical consultation systems. *Comput Biomed Res* 1981;14:542-58.
- Fox J, Barber D, Bardhan KD. Alternative to Bayes? A quantitative comparison with rule-based diagnostic inference. *Methods Inf Med* 1980;19:210-5.
- Fryback DG. Bayes' theorem and conditional non-independence of data in medical diagnosis. *Comput Biomed Res* 1978;11:423-34.
- Zoltie N, Horrocks JC, De Dombal FT. Computer-assisted diagnosis of dyspepsia—report on transferability of a system, with emphasis on early diagnosis of gastric cancer. *Methods Inf Med* 1977;16:89-92.
- Macartney FJ, Rees PG, Daly K, et al. Angiocardiographic appearances of atrioventricular defects with particular reference to distinction of ostium primum atrial septal defect from common atrioventricular orifice. *Br Heart J* 1979;42:640-56.
- Lachenbruch PA. Evaluating a discriminant function. In: Lachenbruch PA, ed. *Discriminant analysis*. New York: Macmillan, 1975:25-39.
- Habbema JDF, Gelpke GJ. A computer program for selection of variables in diagnostic and prognostic problems. *Comput Programs Biomed* 1981;13:251-70.
- Croft DJ, Machol RE. Mathematical methods in medical diagnosis. *Ann Biomed Eng* 1974;2: 69-89.
- Nordyke RA, Kulikowski CA, Kulikowski CW. A comparison of methods for the automated diagnosis of thyroid dysfunction. *Comput Biomed Res* 1971;4:374-89.
- Fleiss JL, Spitzer RL, Cohen J, Endicott J. Three computer diagnosis methods compared. *Arch Gen Psychiatry* 1972;27:643-9.