J. Roy. Coll. Phycns Lond.

Computer-guided Diagnosis

T. R. TAYLOR, MB, MRCP, DPM, Department of Medicine in relation to Mathematics and Computing, University of Glasgow

This paper describes a practical demonstration of the technique of computer-guided diagnosis which Card (1970) calls a Mark 2 model as compared to his Mark 1 model which calculates the probability only when all the available data have been elicited. In clinical practice, the doctor collects data from a patient 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. The sequential diagnostic technique used in this study uses Bayes' Theorem to calculate the probabilities, and selects from the remaining tests the one that is expected to be the most informative in the patient being examined. Technically this is done by selecting the test with the greatest expected informativeness (Lindley, 1956; Good, 1968). In this context the term test includes an item from the history or physical examination as well as any laboratory investigations. Figure 1 shows a list of the tests used in this study.

HISTORY	EXAMINATION
3 – Discomfort	6 – Fixation to tissues
9 – Pain in goitre	7 – Cervical lymph glands
10 – Hoarseness	8 – Pyramidal lobe
11 – Dysphagia	16 – Nodular or diffuse
12 – Choking or tightness	25 – Consistency
13 – Cough or stridor	26 – Clinical status
15 – Recent increase in size	24 – Estimated size of gland
30 – Age	
23 – Duration	
SPECIAL EXAMINATION	
5 – Laryngeal palsy (indirect laryngoscopy)	
4 – Tracheal deviation (or compression on X-ray)	
SERUM TESTS	SPECIAL TESTS
1 – Precipitin test	21 – $PB^{127}I$
2 – Serum globulins	17 – $B.E.^{131}I.$
18 – Gammaglobulin	14 – $KClO_4$ discharge.
19 – ESR	20 – 24 hour thyroidal $^{131}I.$
27 – CF Test	22 – 48 hours $PB^{131}I$
28 – Thymol turbidity	
29 – Zinc sulphate turbidity	

Fig. 1. Tests used in study.

It is important to note that, for the purposes of this study, statistical independence is assumed between the test outcomes (i.e. indicants). This was necessary since the original sample (on which the likelihoods, Fig. 2,

No.	Test	Response	Hashimoto's disease	Simple goitre	Thyroid cancer
25	Consistency	Firm	0·9057	0·5800	0·4600
		Hard	0·0566	0·0400	0·5300
		Soft	0·0377	0·3800	0·0100
27	CF test	++	0·8372	0·0100	0·0200
		+	0·0698	0·0513	0·1081
		−	0·0930	0·9387	0·8719
8	Pyramidal lobe	Absent	0·8491	0·9608	0·9783
		Present	0·1509	0·0392	0·0217

Fig. 2. Example of likelihoods.

were based) was too small to measure their interdependence and to use it in the calculations.

The prior probabilities (i.e. the probability of each disease being seen in the population being studied) were set at equiprobability (i.e. 0·33, 0·33, 0·33) in this series. This was the first of three studies, with varying prior probabilities, conducted to compare the effect of prior probabilities on the results of the sequential technique. The second study has used the prior probabilities of the original clinic population from which the data was collected (i.e. 0·10, 0·89, 0·01), while the third study uses the probabilities corresponding to the 60 cases used in the 3 studies; namely, 34 cases of Hashimoto's disease, 19 of simple goitre, 7 of thyroid cancer, or 0·566, 0·317, 0·107.

DATA

The data used in these calculations are derived from that of Boyle and his co-workers in Glasgow (1966) from a survey of non-toxic goitre in two thyroid clinics. The data used is in the form of likelihoods (Fig. 2), which represent the frequency of test outcomes for each disease. In this table, tests 25 and 27 are highly discriminating tests, while test 8 has similar outcomes for each disease and therefore is not of value within the group of diseases. The 60 cases used to test the Mark 2 model were derived from the same clinics as those used by Boyle *et al.* (1966). In each case the diagnosis was either established histologically (in about 70 per cent of cases) or by agreement among physicians well experienced in thyroid diseases.

METHOD

An 'on-line' link to a computer situated some miles from the department is used. A teletype terminal (Fig. 3) is connected to the computer by a switching device, called a GPO modem, over the public telephone circuit. Access to the computer is obtained by dialling a number using an ex-directory line.

Fig. 3. Teletype and GPO Modem.

When the terminal is in use, the program selects, after a command signal from the teletype keyboard, the 'best' test from a total of 30 (*see* Fig. 1). The terminal prints out the name of the test and awaits entry by the physician of the patient's response, i.e. test outcome, before proceeding to select the next most informative test.

RESULTS

The results of a study of 60 cases are shown in Fig. 4. It shows an overall success rate of 93 per cent with four misdiagnoses out of 60: it is important to note that all cases of thyroid cancer were correctly diagnosed. Two simple goitres were misdiagnosed, one as cancer and the other as Hashimoto's disease. Two cases of Hashimoto's disease were misdiagnosed as cancer.

In the case of simple goitre misdiagnosed as cancer, the thyroid gland was hard and nodular with a recent increase in size and, presumably, went to

	Correct	Wrong	Misdiagnosed as: H	S	C
Hashimoto's disease	32	2		0	2
Simple goitre	17	2	1		1
Thyroid cancer	7	0	0	0	
Total	56	4		=	93%

(H) Hashimoto's disease (S) Simple goitre (C) Thyroid cancer

Fig. 4. Results of study of computer-guided diagnosis.

histology to exclude cancer. 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, raised the very important question of dual pathology, i.e. the co-existence of simple goitre and a mild degree of autoimmune thyroiditis, which is well documented in the literature (Gribetz *et al.*, 1954).

Review of the two cases of Hashimoto's disease misdiagnosed as cancer showed evidence strongly suggestive of cancer. Both patients were euthyroid, both goitres were hard, and no immunological evidence of Hashimoto's disease was demonstrated.

No attempt was made in this study to compare the success rate of the sequential diagnostic technique with that of the physicians. This has already been treated in some detail by Boyle *et al.* (1966). The accuracy in this study was calculated with reference to the final diagnosis established either histologically or by agreement among physicians well experienced in thyroid disease.

Figure 5 shows the analysis of the cases in terms of the number of tests needed to reach a probability of 0·99 or over. Also shown are those cases where the diagnoses were correct but where they failed to reach a final

No. of tests	FINAL PROBABILITY of > 0·99 < 2	3–6	7–10	> 11	< 0·99
Hashimoto's disease	12	12	3	4	1 (0·804)
Simple goitre	0	3	7	3	4 (0·652) (0·902) (0·900) (0·802)
Thyroid cancer	0	5	1	1	0
Total	12	20	11	8	5

Fig. 5. Correct diagnosis.

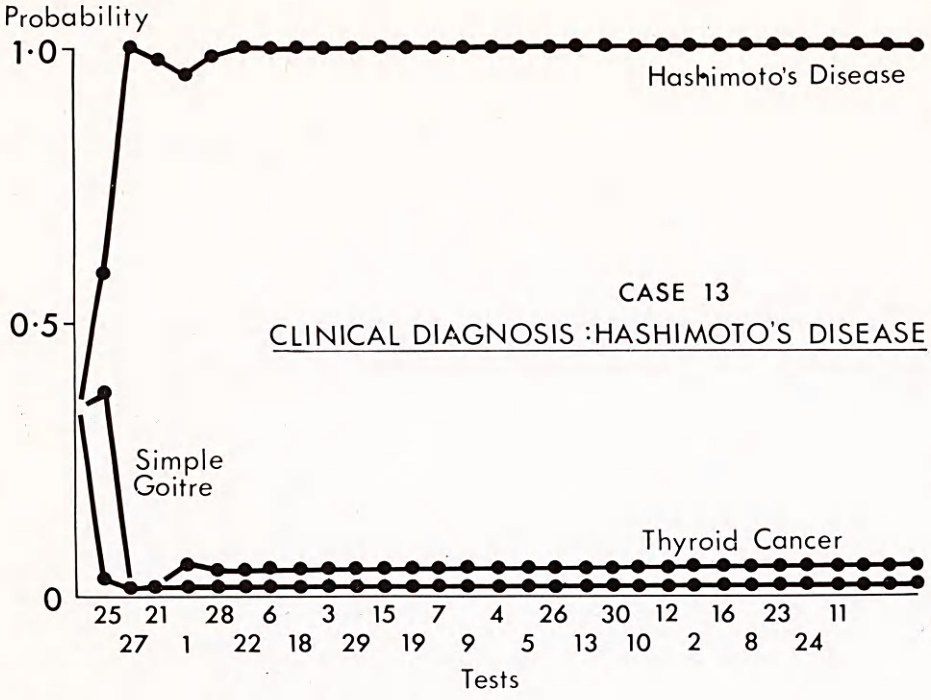

Fig. 6. Graph showing probabilities and tests.

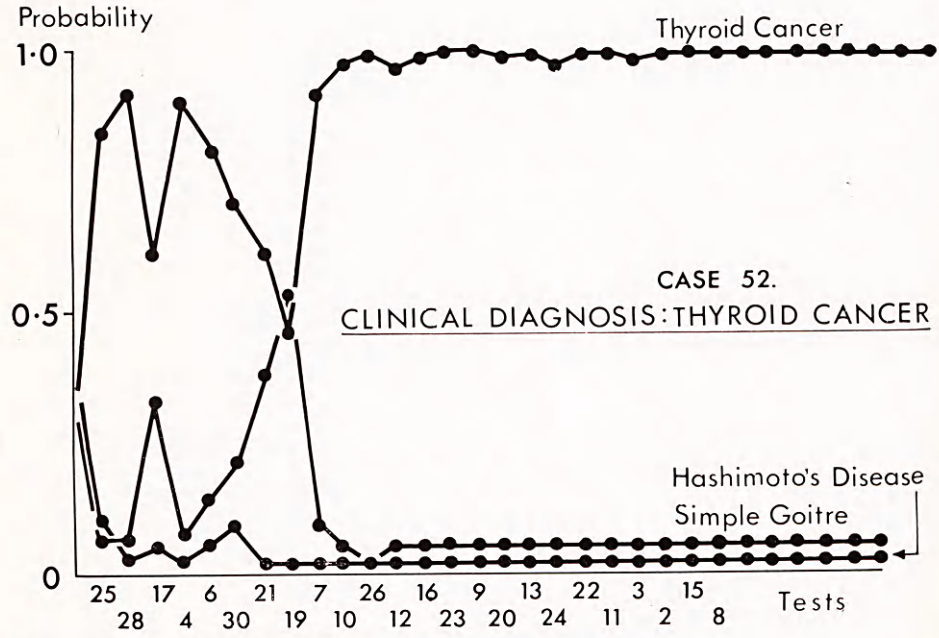

Fig. 7. Illustrative case showing the effect of contradictory information.

probability of 0·99 (the final probability after all available tests were used is shown in parentheses).

Of particular interest are the 12 cases of Hashimoto's disease which needed only two tests to reach 0·99. In each case in this first study (with prior probabilities 0·33, 0·33, 0·33) the most informative test was the consistency of the gland (as it was in all 60 cases); in the 12 cases the second test was the complement fixation text (10), the $PB^{127}I$ and the Precipitin Test (1). The most informative test in both the second and third studies (*see* theoretical basis) was 27 (complement fixation test).

Over half the cases (32 out of 56) needed 6 or less tests out of 30; and only 13 needed more than 10 tests.

ILLUSTRATIVE CASES

The value of a sequential approach is seen best when the changes in probabilities (as tests are selected) are displayed graphically. Figure 6 shows a typical case which took less than seven tests to reach a probability of 0·99. Figure 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.	$PB^{127}I$	$< 3{\cdot}0\ \mu g/100$ ml
19.	ESR	0–20 mm Hg in 1st hour
7.	Cervical lymph glands	Palpable
10.	Hoarseness	Yes

The technique allows us to pinpoint the poor or contradictory items of evidence, i.e. tests 4, 6, 30, 21, 19, and demonstrates how the positive evidence of tests 7 and 10 overcomes this.

DISCUSSION

A number of observations can be made from the results of this study. There is almost certainly a great deal of redundancy in data recorded on patients. Figure 7, for example, demonstrates clearly that the great majority of the tests in this case were superfluous. Figure 1 shows tests under the heading 'serum tests'; of these 1, 28, and 27 are much more discriminating than the others. 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) which show that its rate of occurrence is almost identical in each disease.

The variation among tests in their discrimination is also revealed by the graphs (Figs 6 and 7) 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 erythrocyte sedimentation rate into three classes 0–20, 21–40, and >40 mm in 1 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 particular group of patients, rather than to make a more precise measure of a physiological variable.

The discriminant power of clinical (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 this study 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 discriminant power in this group of diseases; in the second and third studies (*see* theoretical basis) it 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 hours thyroidal ^{131}I uptake and the 48 hours $PB^{131}I$.

Another important aspect of this approach to clinical diagnosis is that an objective level of final diagnosis (i.e. 0·99) has been chosen arbitrarily. It is now open to dispute whether this level, representing an error rate of 1 in 100 cases, is acceptable.

A public theoretical approach to diagnostic medicine, which so far lacks a scientific base, may now be possible. Statistical decision theory may well be a central component of such a science of clinical medicine.

Finally, this study has shown that the use of an on-line computer terminal and such programs as are used in this study makes the use of a computer in diagnosis at least technically feasible in an outpatient clinic. It is, however, important to note that this simple model involved only 30 tests in three diseases. The time delay between entering a patient's response, i.e. test outcome at the terminal, and printing out the next 'best' test, is of the order of $5\frac{1}{2}$–6 seconds. In a much larger model of all thyroid diseases (with 10 or

more diseases and over 100 tests) this time delay may be much longer. The computer used in the present study was relatively small, and it may well be that this problem would be solved by using the much larger time-sharing computers now available.

Acknowledgements

I wish to thank Professor W. I. Card and Professor J. Aitchison for their critical comments, and Dr J. A. Boyle and his co-workers for access to case material. This work is supported by a grant from the Medical Research Council.

This article is based on a paper read at the Conference on Computers in Medicine held at the Royal College of Physicians in September, 1969.

References

Boyle, J. A., Greig, W. R., Buchanan, W. W., Harden, R. McG., Franklin, D. A., McGirr, E. M. (1966) *Quart. J. Med.*, **35**, 565.
Card, W. I. (1970) *J. Roy. Coll. Phycns Lond.*, **4**, 183.
Good, I. J. (1968) *Virginia J. Sci.*, **19**, 101.
Gribetz, D., Talbot, W. B., Crawford, J. D. (1954) *New Engl. J. Med.*, **250**, 555.
Lindley, D. V. (1956) *Ann. Mathematical Statist.*, **27**, 986.

Decision-making in Clinical Medicine

JOHN AITCHISON, MA, FRSEd., Titular Professor in Statistics, University of Glasgow

For centuries physicians have been faced daily with difficult problems of decision-making under uncertainty and have found, in the practice of medicine, resolutions of these problems. The science of decision-making under uncertainty—statistical decision theory—is barely twenty years old (Wald, 1950). Have medical practitioners yet considered what advantages might accrue from their taking the younger discipline into partnership? In such a partnership, what role could the novice best play to ease the burden of medical practice? It is against the background of such general questions that this article examines one particular aspect of decision-making in medicine—the problem of treatment allocation.

While the application of statistical decision theory to real problems in