MEDICAL PRACTICE

Hospital Topics

Use of Sequential Bayesian Model in Diagnosis of Jaundice by Computer

R. P. KNILL-JONES, R. B. STERN, D. H. GIRMES, J. D. MAXWELL, R. P. H. THOMPSON, ROGER WILLIAMS

British Medical Journal, 1973, 1, 530–533

Summary

A sequential Bayesian model has been developed for a computer and used to diagnose jaundiced patients admitted to hospital. Up to 102 items of information from the history, physical examination, and special investigations available within 48 hours of admission were collected on 309 patients. The results from these patients were used to calculate the probabilities of 11 possible diseases in 65 new patients and also to place patients into groups for medical or surgical treatment.

The overall accuracy of the model in diagnosing patients as having one of 11 diseases was 69%, and where the final probability reached > 0·96, it was 89%. The overall accuracy in making a medical or surgical decision was 89%, and where the final probability reached > 0·96 it was 94%.

Improvement in accuracy should result as the number of cases seen with rare conditions increases, and probably a similar model could be developed and used to make most use of those indicants with the highest cost-effectiveness.

Liver Unit, King's College Hospital, London SE5 9RS
R. B. STERN, M.B., M.R.C.P., Research Fellow
J. D. MAXWELL, M.B., M.R.C.P., Research Fellow
R. P. H. THOMPSON, D.M., M.R.C.P., Honorary Senior Lecturer
ROGER WILLIAMS, M.D., F.R.C.P., Consultant Physician and Director

Department of Medicine in Relation to Mathematics and Computing, University of Glasgow, Glasgow W2
R. P. KNILL-JONES, M.SC., M.R.C.P., Research Fellow

Department of Statistics and Computer Science, University College, London WC1
D. H. GIRMES, B.SC., D.I.C., Lecturer in Statistics and Computer Science

Introduction

The clinician concerned with the management of a jaundiced patient may make decisions at two levels. At one level he decides the disease from which the patient is suffering, and at the second level whether the patient requires medical or surgical treatment. He uses the available clinical and laboratory information to reach these decisions as quickly and accurately as possible, but this may be difficult because of apparently conflicting evidence about the disease and because of difficulty in putting the correct weight on the results of investigations. The less certain the decisions the longer the patient stays in hospital, and this increases both the financial and human costs of his management. A formal mathematical approach to this problem might therefore not only improve the speed and accuracy of diagnosis but, by optimal use of the available evidence, also reduce the cost of reaching the diagnosis.

In 1967 we decided to investigate a Bayesian model to study the increasing number of jaundiced patients being referred to the liver unit.[1] Later a sequential method of using the available information about each patient was introduced in an attempt to eliminate unnecessary information. Only information available within 48 hours of admission was considered so that a diagnosis could be obtained as rapidly as possible. The model has also recently been extended to diagnose patients with jaundice in the South-east Metropolitan Region and in one hospital in Glasgow.

Development of Model

It is useful to have a single word to describe each item of information derived from the history, examination, or special investigations and "indicant"[2] is used in this paper. Indicants may be discrete variables—for example, the presence or

absence of abdominal pain or itching—or continuous variables —for example, serum aminotransferases or liver size measured on the ^{99m}Tc scintiscan. The model is based on Bayes' theorem,[3] which in the medical context enables the known frequency of indicants occurring in a disease to be used to calculate the probability of that disease occurring in a new patient with the same indicants.

"PAST EXPERIENCE" PATIENTS

A Bayesian model requires a past experience of patients to provide the probability with which each indicant occurs in each disease in the population under study. Data were therefore collected prospectively from all patients who presented with jaundice from January 1968. These data were updated every time 50 further patients were seen, so that by January 1972 there were 320 patients to provide experience for the model.

The information from each patient, consisting of 45 indicants from the history, 12 physical signs, 34 laboratory results, and, in 206 patients, 11 indicants from the ^{99m}Tc liver scintiscan, was entered on a standard form when the patient had been in hospital for 48 hours. More complex investigations such as immunological tests or liver biopsy were not used by the present model to make a diagnosis since these were not available within 48 hours of admission, and the aim was to achieve an early diagnosis by the computer model.

The data from the past experience group of patients were analysed by a program written in FORTRAN for the University of Edinburgh I.B.M. 370/155 computer, which is linked by telephone to the computer department of the University of Glasgow. The results of continuous variables were summarized (with logarithmic transformation where necessary) as the mean and standard deviation for each disease. The value of a continuous variable for a new patient was compared by the computer with the known distribution of the results of that variable for each disease in turn, and the probability of that value occurring in each disease was then calculated using a modification of Bayes's theorem.[4] It was assumed that indicants for a given disease were independent of one another.[5]

The final clinical diagnosis on each patient was made from the combined clinical and laboratory data, with the results of liver biopsy, laparotomy, necropsy, or follow-up. In addition to the patients forming the past experience group a further 11 patients had been seen, but these were withdrawn from the model either because their diagnoses never became certain or because they were too rare to be included at present. The remaining 309 patients were diagnosed as having one of 11 diseases (see table). The term "acute viral hepatitis" was used to include both infectious and serum hepatitis. The criteria for "fulminant hepatic failure" were as defined by Trey and Davidson.[6] "Cirrhosis" did not include patients with active chronic hepatitis or primary biliary cirrhosis, who had separate disease categories. "Infiltration of the liver" included primary hepatoma, metastatic carcinoma, and reticulosis. "Drug-induced jaundice" was divided into that due to chlorpromazine hypersenitivity and that due to other agents.

Distribution of Patients among 11 Disease Groups

Disease Group	"Past Experience" Patients		New Patients	
	No.	%	No.	%
Acute viral hepatitis	56	18·1	14	21·5
Fulminant hepatic failure	25	8·1	9	13·8
Cirrhosis	25	8·1	9	13·8
Active chronic hepatitis	25	8·1	6	9·2
Primary biliary cirrhosis	18	5·8	1	1·6
Infiltration of liver by tumour	14	4·5	1	1·6
Chlorpromazine jaundice	12	3·9	2	3·1
Other drugs (see text)	11	3·6	2	3·1
Gilbert's disease	9	2·9	1	1·5
Total "medical"	195	63·1	45	69·2
Tumour of extrahepatic biliary tree	79	25·6	18	27·7
Gall stones	35	11·3	2	3·1
Total "surgical"	114	36·9	20	30·8

USE OF MODEL IN DIAGNOSIS OF NEW PATIENTS

The ability of the model to reach a diagnostic decision was tested in 65 new patients presenting with jaundice at the liver unit, at hospitals in the South-east Metropolitan Region, and at the Southern General Hospital, Glasgow, between January and July, 1972. There was no significant difference in the incidence of each disease between the past experience group and the new patients ($\chi^2 = 11{\cdot}8$, $P > 0{\cdot}2$) (see table).

Two decisions were made for each patient. Firstly, a diagnosis of one of the 11 diseases was made. Secondly, the probability was calculated for a "medical" or "surgical" cause of jaundice. For this the probability of a surgical decision was obtained by adding the probabilities of a tumour of the extrahepatic biliary tree or head of the pancreas and of gall stones, while the sum of the probabilities of the other nine diseases together became that of a medical decision.

Sequential Selection.—Rather than use in a random order all the indicants which had been obtained from each case the model selected sequentially that test which, at each stage, was calculated to give the greatest expected gain in information. This procedure reduced the number of tests needed to come to a diagnosis at a given level of probability. Taking about 10 seconds per patient the computer calculated the probability of each of the 11 diseases and, from these, that of a medical or surgical decision.

In fig. 1 is shown how a diagnosis was reached in a new patient using sequential selection of tests. For clarity the probabilities of only two of the 11 possible diseases are shown. On the ordinate is a probability scale from 0·0 (impossible) to 1·0 (definite). Along the abscissa tests are shown in the order selected by the computer. The previous probability of each disease in the population is first plotted, for this is the incidence of these diseases in the past experience group and will therefore be the same for any new patient until the data are updated. Thus the previous probability of any new patient having extrahepatic obstruction due to a tumour of the biliary tree or the pancreas is 0·26, and of having acute viral hepatitis 0·18 (table). The first test selected is always the serum aspartate aminotransferase, and the computer then uses the value in this patient of 100 IU/l. to calculate the new probabilities of each disease. These proba-

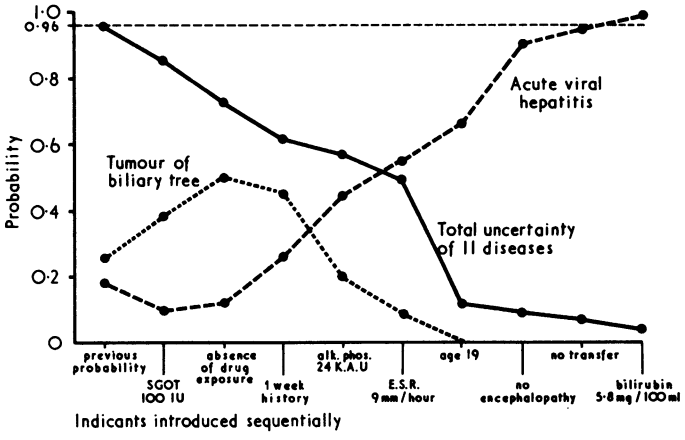

FIG. 1—Sequential probability map of the diagnoses of a patient in whom, using only nine indicants, a diagnosis of viral hepatitis was made at a final level of probability of 0·98. "No transfer" indicates that the patient was admitted direct to King's College Hospital. "Total uncertainty" is a measure of the probabilities of all 11 diseases, scaled from 0 to 1.

bilities will change further as more information about a new patient is introduced. In this example the probability of hepatitis actually fell to 0·09 and that of a tumour of the biliary tree rose to 0·39. The computer then scans through the tests which might be available on a new patient and selects as the next test the one which on the basis of the model's past experience is likely to give the greatest gain in information on that patient; it then uses the value of this test for the new patient to alter the probabilities of each disease. The test selected differs from patient to patient, depending on the results of the previous test. In this patient the one next most likely to lead to a diagnosis was a history of exposure to potentially hepatotoxic drugs, and therefore the computer chose this indicant as the one to introduce next. This patient was not exposed, and this again altered the probabilities of each disease. Thus only the previous probabilities and the first indicant are the same for any new patient, and the order of selection of the other indicants differs from patient to patient, depending on the results of each indicant. In this way indicants are sequentially selected until either a final probability for one disease greater than 0·96 is reached or until all available indicants have been used. In the example the final diagnosis of viral hepatitis was reached with a probability of 0·98 using only nine tests even though the value of the first test initially decreased its probability.

Accuracy of Computer Diagnosis in New Patients

DIAGNOSIS

In all 65 patients the final diagnosis was reached with a probability of 0·35 or greater, but only 36 patients (55·4%) reached a final diagnosis with a probability level of greater than 0·96. The percentage of patients in whom the computer diagnosis was subsequently found to be correct increased with increasing probability of the computed diagnosis (fig. 2). Thus when a computed diagnosis had a probability of 0·4 it was correct in 71% of cases, but when it had a probability of 0·96 it was correct in 89% of these cases.

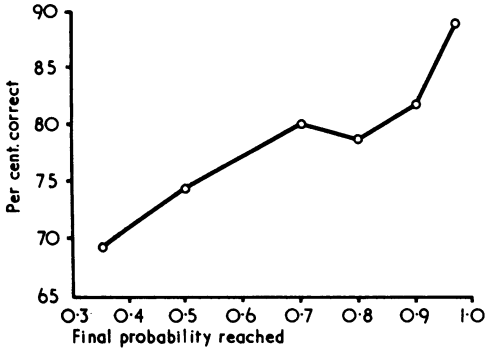

FIG. 2—Percentage of patients correctly diagnosed by the model as between 11 diseases at each final probability level of diagnosis.

In four patients computed diagnoses were made with a probability of > 0·96 and yet were subsequently found to be incorrect. In the first a poor uptake of ^{99m}Tc in the scintiscan favoured a computer diagnosis of cirrhosis, but he was later proved to have viral hepatitis. In the second the presence of rigors favoured a computer diagnosis of gall stones, but he was later proved to have viral hepatitis. In the other two patients in whom computer diagnoses of fulminant hepatic failure were made the error was due only to our definition, for both patients had severe hepatic failure but it was secondary to an underlying cirrhosis.

MEDICAL VERSUS SURGICAL DECISION

In all 65 patients the final diagnosis was reached with a probability of 0·5 or greater. Again fewer patients reached the highest levels of probability, but in 52 patients (80%) a decision was reached with a probability level of greater than 0·96. The percentage of patients in whom the computed decision was subsequently found to be correct also increased with increasing probability of the decision. Thus when the computed decision had a probability of 0·5 it was correct in 89% of cases, and when it had a probability of 0·96 it was correct in 94% of cases.

In three patients the computed decision was made with a probability of > 0·96 but was later found to be incorrect. The first, given a "surgical" decision by the computer, had a long history of weight loss and jaundice and a high serum alkaline phosphatase but was later proved to have viral hepatitis with prolonged cholestasis. The second, given a "medical" decision, was later found to have obstructive jaundice due to a carcinoma of the gall bladder which was invading the liver substance. The third was the patient finally proved to have hepatitis referred to above.

The present model was retested using indicants in a random order, different for each patient, instead of sequentially. The random selection needed a mean of 34 (range 12-77) indicants to reach a diagnosis with a probability of > 0·96, and the sequential selection needed a mean of only 14 (range 4-33). This difference is significant ($P < 0·01$) (Wilcoxon's test for pair differences) and is an indication of the effectiveness of the sequential method.

Discussion

Previous workers have applied Bayesian principles to the diagnosis of liver disease and jaundice. Burbank[7] produced a model for the diagnosis of six "prolonged undifferentiated liver diseases"—gall stones, active chronic hepatitis, primary biliary cirrhosis, drug cholestasis, carcinoma of the biliary tree, and hepatitis with prolonged cholestasis. He achieved a correct diagnosis in 58% and a correct medical or surgical decision in 92%, but unlike our study he also used indicants from the liver biopsy. Our model was therefore also tested in 66 patients from the past experience group and 64 from a new patient group in whom jaundice had persisted for more than four weeks. These patients were thus similar in number and length of history to the patients used by Burbank. The accuracy of the present model under these conditions was 66% for six diseases and 92% for a medical or surgical decision, the results being very similar to those found by Burbank. However, because our model did not use indicants from the liver biopsy a diagnosis was reached more quickly. Bégon and Dhumeaux[8] produced a model using only biochemical indicants, and the percentage of correct answers was 80% for the four diagnoses given by the computer as the most likely out of 35 possible diseases. Vishnevsky *et al.*[9] achieved an 80% success rate in diagnosing 15 different surgical causes of jaundice, and Cattaneo *et al.*[10] achieved 71% accuracy when their model was tested under similar conditions to ours.

We have not attempted to compare our results directly with those of a physician, but our computer diagnosis compares very favourably with the results of other studies. Martin *et al.*[11] found that specialist physicians made a correct medical *v.* surgical prediction in 87% of cases, and Schenker *et al.*[12] reported a figure of 85%. Cattaneo *et al.*[10] found that clinicians were correct in 86% of cases of jaundice due to one of 10 surgical conditions or to a medical cause. Bégon and Dhumeaux[8] found that the diagnostic performance of two specialists in liver disease was almost identical with their computer model. Our results thus show that when making a medical *v.* surgical decision, and when diagnosing one of the 11 diseases (provided a level of probability of > 0·96 is reached), the model reaches a

similar accuracy to these physicians. However, the groups of patients may not have been exactly comparable.

The results of the sequential model are somewhat better than those of our previous non-sequential model using 172 patients,[1] in whom the accuracy of diagnosis as between nine diseases was 43% using ^{99m}Tc scintiscan indicants alone, 58% using laboratory indicants alone, 65% using clinical indicants alone, and only marginally better using all three combined. The figures for a medical *v.* surgical decision were 75%, 80%, and 82% respectively. Our present model achieves an 89% accuracy for a medical *v.* surgical decision.

There are two main reasons for incorrect diagnoses made by the model. Firstly, although more information is obtained from the history than from the physical examination, laboratory tests, or ^{99m}Tc scintiscan, history-taking is the least accurate source of information. This was shown by an observer error study performed by one of us (R.B.S.) on 25 of the new patients in which each patient answered a self-administered questionnaire as well as the standard form. When using indicants from the history alone the diagnoses reached by the model using indicants from the doctor's form and then from the patient's questionnaire agreed in only two of the 25 patients.

Secondly, the past experience of the model is still relatively small, and there are a few atypical cases that have not been seen sufficiently often to be recognized as a separate diagnostic category. For example, two errors were due to the fact that the model regarded a prolonged prothrombin time and small liver on ^{99m}Tc scintiscan as favouring fulminant hepatic failure rather than cirrhosis, and one error was due to regarding a long history as favouring carcinoma of the biliary tree rather than atypical viral hepatitis. These errors will become fewer in the future, since the model's past experience is enlarged and updated every time 50 further new patients are seen. As these numbers increase, both hepatic failure in association with cirrhosis and hepatitis with prolonged cholestasis will be regarded as additional distinct diagnostic categories by the model. Improvement in the accuracy of a Bayesian model with increasing numbers of patients has been shown to occur in the diagnosis of acute abdominal pain,[13] and in the present study accuracy improved from 74% using 220 patients to 89% when 309 patients were available for past experience.

If in the future it could be decided which indicants are most often redundant in reaching a sequential diagnosis a great saving in cost and time could be made in obtaining information. The model is being developed so that given the cost of obtaining each indicant the cost of reaching a diagnosis can be minimized. This has been shown to be possible for thyroid disease.[14]

The help and advice of Professor W. I. Card is gratefully acknowledged. We thank D. Teather for help with statistical methods and Mary Nicholson for computer programming. R.P.K.-J. was in receipt of a grant from the Medical Research Council, R.B.S. a Bernard Sunley Anniversary Research Fellowship from the Royal College of Physicians, and R.P.H.T. a Medical Research Council Clinical Research Fellowship.

References

1 Knill-Jones, R., Maxwell, D. M., Thompson, R. P. H., and Williams, R., *Gut*, 1970, 11, 1062.
2 Card, W. I., and Good, I. J., *Methods of Information in Medicine*, 1971, 10, 176.
3 Bayes, T., *Philosophical Transactions of the Royal Society of London*, 1763, 3, 370.
4 Bailey, N. T. J., *Proceedings of Medical Research Council Conference on Mathematics and Computer Science in Biology*, No. 103. London, H.M.S.O., 1965.
5 Bailey, N. T. J., *Mathematical Approach to Biology and Medicine*, chapt. 11. London, Wiley, 1967.
6 Trey, C., and Davidson, C. S., *Progress in Liver Disease*, ed. H. Popper and F. Schaffner, vol. 3, chapt. 11, p. 282. London, Heinemann, 1970.
7 Burbank, F., *American Journal of Medicine*, 1969, 46, 401.
8 Bégon, F., and Dhumeaux, D., *Minnesota Medicine*, 1971, 54, 101.
9 Vishnevsky, A. A., *et al.*, *Eksperimental'naya Khirurgiya*, 1964, 4, 22.
10 Cattaneo, A. D., Luchelli, P. E., Rocca, E., Mattioli, F., and Becchi, G., *Abdominal Surgery*, 1972, 14, 71.
11 Martin, W. B., Apostalakos, P. C., and Roazon, H., *American Journal of Medical Sciences*, 1960, 240, 571.
12 Schenker, S., Balint, J., and Schiff, L., *American Journal of Digestive Diseases*, 1962, 7, 499.
13 de Dombal, F. T., Leaper, D. J., Staniland, J. R., McCann, A. P., and Horrocks, J., *British Medical Journal*, 1972, 2, 9.
14 Taylor, T. R., Shields, S., and Black, R., *Lancet*, 1972, 2, 79.

Pitfalls in the Diagnosis of Jaundice due to Carcinoma of the Pancreas or Biliary Tree

R. B. STERN, R. P. KNILL-JONES, ROGER WILLIAMS

British Medical Journal, 1973, 1, 533–534

Summary

Analysis of 56 patients with obstructive jaundice due to carcinoma of the pancreas or extrahepatic biliary tree showed that unexpected features were present in 25%. Presentation with painless jaundice was uncommon, and the symptoms were more often non-specific, with malaise, anorexia, and vomiting. Abdominal pain was frequent, and the condition was found in young patients. One-fifth presented with serum alkaline phosphatase levels of less than 30 K.A. units. Some had high serum aspartate aminotransferase levels, more characteristic of hepatocellular jaundice. A mathematical model may be helpful in correctly weighting these various criteria.

King's College Hospital and Medical School, London SE5 9RS
R. B. STERN, M.B., M.R.C.P., Research Fellow, Liver Unit
ROGER WILLIAMS, M.D., F.R.C.P., Consultant Physician and Director of Liver Unit

Department of Medicine in Relation to Mathematics and Computing, University of Glasgow, Glasgow W2
R. P. KNILL-JONES, M.SC., M.R.C.P., Research Fellow

Introduction

During the past five years we have been using a computer-assisted mathematical model based on Bayesian probability theory for the diagnosis of patients with jaundice.[1] Data have been collected from 286 patients, so the validity of accepted diagnostic criteria can be analysed with some certainty. This paper is concerned with certain anomalies of the presenting symptoms, signs, and laboratory investigations of patients with obstructive jaundice due to carcinomata. These have been described before[2-7] but they are not widely appreciated