Alpha₁-fetoprotein in the diagnosis of hepatoma: statistical and cost benefit aspects

P. J. PHILLIPS, R. ROWLAND, D. P. REID, AND M. E. COLES

From the Institute of Medical and Veterinary Science, Frome Road, Adelaide, and the Royal Adelaide Hospital, North Terrace, Adelaide, South Australia 5000

SUMMARY A rational comparison of different serum concentrations of alpha₁-fetoprotein (S-AFP) in the diagnosis of hepatoma must be made. We took data on the sensitivity and specificity of different diagnostic S-AFP concentrations from the literature and evaluated them statistically and by Bayesian analysis. In our patients (hepatoma prevalence 0.028) a sensitive diagnostic concentration (30-50 ng/ml) will misdiagnose hepatoma so often that a positive test will indicate hepatoma in only 10% of cases. A positive test at a specific diagnostic concentration (500 ng/ml) indicates hepatoma in 100% of cases and is preferable in terms of cost benefit. Although the lower concentration will diagnose a larger proportion of patients with hepatoma (74% compared with 59%) the 'costs' of excluding false positives are considerable (A\$2545 per extra case with 2.5% of patients suffering significant morbidity). In western societies, where the prevalence of hepatoma is low, a higher, less sensitive but more specific diagnostic S-AFP concentration is appropriate.

Methodological advances have enabled accurate and reproducible measurements to be made of minute amounts of substances in biological fluids. Unfortunately there is usually an overlap between the diseased and the non-diseased population tested and as an investigation becomes more sensitive¹ it becomes less specific.² The use of serum alpha₁-fetoprotein (S-AFP) determination in the diagnosis of primary hepatocellular carcinoma (hepatoma) illustrates that the choice between sensitivity and specificity can be a calculated one.

Hepatoma often presents a diagnostic problem since its clinical manifestations are protean and it is usually associated with long-standing cirrhosis (Ihde *et al.*, 1974). The definitive diagnosis is histological but S-AFP determination is used in making a presumptive diagnosis. Although patients with hepa-

toma can have abnormal S-AFP concentrations so can patients with a wide variety of hepatic and other conditions (Ruoslahti and Seppälä, 1972a; Purves and Purves, 1972; Chayvialle and Ganguli, 1973). There is obviously an optimal diagnostic concentration which will identify most hepatoma patients and not misdiagnose too many non-hepatoma patients, but it has not been defined (Kew, 1974).

This uncertainty can be partly resolved by statistical analyses. By plotting the true positive (TP) ratio¹ against the false positive (FP) ratio² a curve called the 'receiver operating-characteristic' (ROC) can be constructed (McNeil *et al.*, 1975) (Fig. 1). The curve joins (0,0) the most specific position (where no diseased but all normal patients are correctly diagnosed) to (1,1) the most sensitive position (where all diseased but no normal patients are correctly diagnosed). The optimal operating position lies somewhere between these two points and occurs where the slope of the ROC curve equals (Swets *et al.*, 1964)

$$\frac{\text{additional cost of false positive result} \times \text{probability of no disease}}{\text{additional cost of false negative result} \times \text{probability of disease}}$$

This complex formula is obvious in general terms. If a disease is common and treatment is curative and harmless to the normal patient a sensitive test (or diagnostic concentration) is appropriate and one uses a position to the right where the slope is low. If the disease is rare and the treatment is only palliative

Correspondence should be addressed to P. J. Phillips, Division of Clinical Chemistry, Institute of Medical and Veterinary Science, Frome Road, Adelaide, South Australia 5000.

Received for publication 28 April 1977

¹A sensitive test is one which correctly identifies diseased patients. It is positive in diseased patients and has a high true positive ratio.

²A specific test is one which correctly identifies non-diseased patients. It is negative in non-diseased patients and has a low false positive ratio (Galen and Gambino, 1976).

Fig. 1 ROC curve for different diagnostic serum AFP concentrations. Numbers refer to references from which the data were taken (see Table 1).

Table 1 References for Fig. 1

Point in Fig. 1	Source of data
1	Abelev (1971)
2	Bloomer <i>et al.</i> (1975)
3	Chayvialle <i>et al.</i> (1974)
4	Elgort <i>et al.</i> (1972)
5	Foli <i>et al.</i> (1969)
6	Kohn and Weaver (1974)
7	Muñoz <i>et al.</i> (1972)
8	Ruoslahti and Seppälä (1972b)
9	Smith and O'Neill (1971)
10	Teres <i>et al.</i> (1970)
11	Zawadski and Kraj (1974)

and dangerous to the normal patient a specific test is appropriate and one uses a position to the left where the slope is high.

The ROC curve can also be used to determine the additional information gained by performing the test (Metz *et al.*, 1973). The ROC curve for guessing is a 45° line through the origin and any test which has TP and FP ratios giving a point above this line reduces uncertainty. For a given disease prevalence the extra information gained is maximal at one particular point (the Imax point). This point may be the appropriate operating position.

Tests or diagnostic concentrations with different sensitivities and specificities can also be evaluated by determining the relative 'cost' of a positive diagnosis. 'Costs', however, are difficult to define in purely financial terms and the extra 'cost' of using a more sensitive diagnostic strategy may be justified on

non-financial grounds.

Bayesian analysis gives the probability that a positive test in a given population indicates disease and is a further technique of comparison.

This study uses these three approaches in an attempt to help the chemical pathologist to choose the diagnostic S-AFP concentration which is neither too specific nor too sensitive but 'just right'.

Material and methods

LOCAL EXPERIENCE

Over a five-year period 29 patients from the Royal Adelaide Hospital were diagnosed as having a hepatoma. The histological examinations and S-AFP measurements were done at the Institute of Medical and Veterinary Science. The diagnosis was made histologically in 27 patients and from the clinical picture, hepatic scintigraphy, and arteriography in two. S-AFP was measured by an immunodiffusion technique (limit of detection 500 ng/ml) (Kohn, 1970) with 10 positive (range 500-70 000 ng/ml) and seven negative results (TP ratio = 0.59). There were no false positive results in 581 other patients, most of whom had been diagnosed as having cirrhosis.

STATISTICAL ANALYSES

Using the ROC curve

An ROC curve (Fig. 1) was constructed from data in reports of different assay methods for AFP used in patients who had liver disease and were of predominantly western European or similar extraction (a total of 543 hepatoma patients and 4362 other patients had been studied). The numbers in Fig. 1 refer to the references from which the data were taken (Table 1). Additional costs are difficult to establish since they include those of performing inappropriate investigations and giving inappropriate treatment. Since hepatoma has a poor prognosis and is not generally amenable to treatment in the group of patients considered (Linder *et al.*, 1974) we may reasonably assume that the additional cost of a false positive (misdiagnosing a patient without hepatoma) is greater than that of a false negative (misdiagnosing a patient with hepatoma). The position of the optimal diagnostic S-AFP concentration will then have a slope greater than

$$\frac{\text{Probability of no hepatoma (PH-)}}{\text{Probability of hepatoma (PH+)}}$$

The points where $\frac{\text{PH-}}{\text{PH+}}$ equals the slope of the ROC

curve are shown for the observed (17/598, 0.028) and two higher hepatoma prevalences (0.1, 0.2) in Fig. 2. The ROC curve was also used to determine the

Fig. 2 Optimal operating positions and Imax points for different hepatoma prevalences. A is the position of the diagnostic serum AFP concentration of the present study (500 ng/ml; TP ratio = 0.59, FP ratio = 0). B is the position of a more sensitive diagnostic serum AFP concentration (50-100 ng/ml; TP ratio = 0.93, FP ratio = 0.20).

points where the additional information gained by the test was maximal (the Imax points) for these three hepatoma prevalences (Fig. 2).¹

COSTING

The optimal operating position determined by the cost benefit analysis using the ROC curve takes the 'costs' of false positive and negative results into account. Alternatively, two diagnostic S-AFP concentrations may be compared by determining how much it 'costs' to diagnose an extra case by the more sensitive investigative strategy.

For simplicity, only two diagnostic S-AFP concentrations were considered: this study's less sensitive, more specific concentration A and a more sensitive, less specific concentration B determined by radioimmunoassay with characteristics determined from the ROC curve (50-100 ng/ml; TP ratio = 0.93, FP ratio = 0.20), (Ruoslahti and Seppälä, 1972b; Chayvialle *et al.*, 1974).

We assumed that if the patient's S-AFP concentration was greater than A in the appropriate clinical

setting the diagnosis would be considered established and no further investigations would be performed. If the patient's S-AFP concentration was less than A but greater than B a liver scan would be performed. If this were positive a scan-guided percutaneous liver biopsy would be performed to establish the diagnosis. If the patient's S-AFP concentration was less than B hepatoma would be considered to be excluded. Table 2 gives the data concerning the investigations which were used in the calculation of costs for the two diagnostic S-AFP concentrations: the TP and FP ratios, morbidity, and mortality were taken from the literature (Ludbrook *et al.*, 1972; Conn, 1972; Zamcheck and Klausenstock, 1953; Terry, 1952) and the costs quoted are local ones.

By Bayesian analysis

The probability of hepatoma in a patient whose S-AFP concentration was positive or negative at concentration A or concentration B was calculated by Bayesian analysis using the TP and FP ratios and three prevalences quoted above.

Results

Fig. 2 shows that the Imax points and the positions of the optimal diagnostic S-AFP concentrations for the three prevalences are closer to the position of A than to the position of B. The superiority of the less sensitive diagnostic S-AFP concentration is most marked at a low prevalence hepatoma.

Table 3 shows the yield and cost per hepatoma diagnosed by using the less sensitive diagnostic concentration (A) or the more sensitive diagnostic concentration (B) followed by confirmatory investigations. The marginal cost of the more sensitive investigative procedure (the cost of diagnosing each case not diagnosed by using diagnostic concentration A alone) is large (\$A2545 for the hepatoma prevalence of this series) and there is significant morbidity.

Table 2 Data concerning investigations

Technique	Diagnostic AFP concentration A	Diagnostic AFP concentration B	Liver* scan	Liver biopsy
TP ratio	0.59	0.93	0.64	0.70†
FP ratio	0	0.20	0.08	0
Cost (A\$)	10	10	39	133‡
Morbidity (%)	0	0	0	0.5§

*Ludbrook *et al.* (1972).

†Conn (1972).

‡Cost of liver biopsy includes local fees for a 24-hour hospital stay, the biopsy procedure, clotting studies, and blood grouping and matching.

§Mortality 0.17% (Zamcheck and Klausenstock, 1953). Morbidity 0.32% (Terry, 1952).

¹We would be pleased to supply interested readers with tables showing the source of data used in constructing the ROC curve, the details of the costing structure used, and the co-ordinates of the Imax points and optimal operating positions.

Table 3 *Costs of diagnosis of hepatoma (to nearest \$A5)*

<i>Hepatoma prevalence</i>	<i>0.028</i>		<i>0.1</i>		<i>0.2</i>	
	<i>A</i>	<i>B*</i>	<i>A</i>	<i>B*</i>	<i>A</i>	<i>B*</i>
<i>Diagnostic AFP concentration used</i>						
Yield (%)†	59	74	59	74	59	74
Average cost (A\$)	605	1000	170	310	85	180
Marginal cost (A\$ per extra case)	—	2545	—	860	—	540
Morbidity (per 100 extra cases)‡	—	2.5	—	1.2	—	0.9

*Costs for B include those of necessary confirmatory investigations.

†Yield is percentage of hepatomas in population that are diagnosed.

‡Morbidity is the number of patients suffering investigative ill effects per 100 extra cases diagnosed by using diagnostic concentration B and confirmatory investigations but missed by using diagnostic concentration A alone.

Table 4 shows that if the S-AFP concentration is A for the hepatoma prevalence of this series (0.028) hepatoma will be present in 100% of cases, whereas if it is B hepatoma will be present in only 12% of cases.

Discussion

Hepatoma is uncommon in western societies and may be suspected in any patient with deteriorating cirrhosis. If S-AFP is measured to diagnose hepatoma our analyses show that the routine use of a more specific, less sensitive diagnostic S-AFP concentration is appropriate. We do not have enough data on the use of S-AFP measurement for diagnosing hepatoma in other groups of patients. A therapeutic enthusiast might think that measurement of S-AFP concentrations might be useful in detecting hepatoma in patients at risk when the disease is pre-clinical and treatment might be more effective. Given the likely hepatoma prevalence and TP and FP ratios, however, such screening would mean a large number of patients undergoing further investigations for a small yield.

Table 4 *Hepatoma probability for different diagnostic serum AFP concentrations and hepatoma prevalences*

<i>Hepatoma prevalence</i>	<i>0.028</i>		<i>0.10</i>		<i>0.20</i>	
	<i>A</i>	<i>B</i>	<i>A</i>	<i>B</i>	<i>A</i>	<i>B</i>
<i>Diagnostic AFP concentration used*</i>						
Probability of hepatoma if test positive	1	0.12	1	0.34	1	0.54
Probability of hepatoma if test negative	0.01	0.003	0.04	0.01	0.09	0.02

*A = less sensitive, more specific diagnostic serum AFP concentration (500 ng/ml; TP ratio = 0.59, FP ratio = 0).

B = more sensitive, less specific diagnostic serum AFP concentration (50-100 ng/ml; TP ratio = 0.93, FP ratio = 0.20).

The results of our analyses in cirrhotic patients are affected by the probability of hepatoma before measurement of S-AFP. Although the usual hepatoma probability in tested patients is low (< 0.03) there are occasional patients in whom the clinician strongly suspects a hepatoma and the probability may be higher (for example, 0.05-0.5). Theoretically there is an optimal operating position on the ROC curve of a test (S-AFP determination) for a particular prevalence of disease (hepatoma) and that point could be used in diagnosis. An approach to the ideal is possible if an ROC curve is constructed and if the disease probability in the individual patient is estimated by a scoring system of clinical and other features. An operating position (diagnostic S-AFP concentration) can be selected to maximise information or minimise costs. The laboratory can then report the optimal diagnostic serum concentration with the quantitative result or state that the result was positive or negative at this concentration. Alternatively the probability of a particular test result (S-AFP concentration) indicating disease (hepatoma) in an individual patient can be calculated and reported with the result. This is easily done by using the ROC curve to determine the FP and TP ratios associated with the patient's result, the clinician's estimate of the patient's disease probability, and Bayes's formula. The appropriate operating position on the ROC curve and the disease probability are easily calculated by a computer and would aid in interpreting the test result.

We used statistical methods to develop a rational approach to the use of a test (AFP determination) in the diagnosis of a disease (hepatoma). Although some details are peculiar to the locality the methods are generally applicable. For hepatoma, whatever the method of measuring S-AFP, a less sensitive, more specific diagnostic concentration (500 ng/ml) should be used for the usual western population and a more sensitive, less specific diagnostic concentration (50-100 ng/ml) should be used only for special patient populations with a high prevalence of hepatoma.

We thank the visiting staff of the Royal Adelaide Hospital for permission to study their patients and Dr R. W. Pain for constructive criticism and advice.

References

- Abelev, G. I. (1971). Alpha-fetoprotein in ontogenesis and its association with malignant tumors. *Advances in Cancer Research*, **14**, 295-358.
- Bloomer, J. R., Waldmann, T. A., McIntire, K. R., and Klatskin, G. (1975). α -Fetoprotein in non-neoplastic

- hepatic disorders. *Journal of the American Medical Association*, **233**, 38-41.
- Chayvialle, J. A. P., and Ganguli, P. C. (1973). Radioimmunoassay of alpha-fetoprotein in human plasma. *Lancet*, **1**, 1355-1357.
- Chayvialle, J. A. P., Touillon, C., Crozier, C., and Lambert, R. (1974). Radioimmunoassay of α -fetoprotein in human serum. *American Journal of Digestive Diseases*, **19**, 1102-1110.
- Conn, H. O. (1972). Percutaneous versus peritoneoscopic liver biopsy. *Gastroenterology*, **63**, 1074-1075.
- Elgort, D. A., Abelev, G. I., and O'Connor, G. T. (1972). Dependence of the specificity of the serologic test for primary liver cancer in different areas of the world on sensitivity of the method used for detecting alpha-fetoprotein. *International Journal of Cancer*, **10**, 331-337.
- Foli, A. K., Sherlock, S., and Adinolfi, M. (1969). Serum α_1 -fetoprotein in patients with liver disease. *Lancet*, **2**, 1267-1269.
- Galen, R. S., and Gambino, S. R. (1976). *Beyond Normality: the Predictive Value and Efficiency of Medical Diagnosis*. Wiley, New York.
- Ihde, D. C., Sherlock, P., Winawer, S. J., and Fortner, J. G. (1974). Clinical manifestations of hepatoma. *American Journal of Medicine*, **56**, 83-91.
- Kew, M. (1974). Alpha-fetoprotein in primary liver cancer and other diseases. *Gut*, **15**, 814-821.
- Kohn, J. (1970). Method for the detection and identification of alpha-fetoprotein in serum. *Journal of Clinical Pathology*, **23**, 733-735.
- Kohn, J., and Weaver, P. C. (1974). Serum-alpha₁-fetoprotein in hepatocellular carcinoma. *Lancet*, **2**, 334-337.
- Linder, G. T., Crook, J. N., and Cohn, I., Jr. (1974). Primary liver carcinoma. *Cancer*, **33**, 1624-1629.
- Ludbrook, J., Slavotinek, A. H., and Ronai, P. M. (1972). Observer error in reporting on liver scans for space-occupying lesions. *Gastroenterology*, **62**, 1013-1019.
- McNeil, B. J., Keeler, E., and Adelstein, S. J. (1975). Primer on certain elements of medical decision making. *New England Journal of Medicine*, **293**, 211-215.
- Metz, C. E., Goodenough, D. J., and Rossmann, K. (1973). Evaluation of receiver operating characteristic curve data in terms of information theory, with applications in radiography. *Radiology*, **109**, 297-304.
- Muñoz, J. R., Guillen, G., and Toribio, F. P. (1972). Alpha₁-fetoprotein in hepatomas, neoplasms, and other illnesses (letter). *Lancet*, **1**, 850-851.
- Purves, L. R., and Purves, M. (1972). Serum alpha-fetoprotein. VI. The radio-immunoassay evidence for the presence of AFP in the serum of normal people and during pregnancy. *South African Medical Journal*, **46**, 1290-1297.
- Ruuslahti, E., and Seppälä, M. (1972a). α -Foetoprotein in normal human serum. *Nature (London)*, **235**, 161-162.
- Ruuslahti, E., and Seppälä, M. (1972b). Normal and increased alpha₁-fetoprotein in neoplastic and non-neoplastic liver disease (letter). *Lancet*, **2**, 278-279.
- Smith, J. B., and O'Neill, R. T. (1971). Alpha-fetoprotein occurrence in germinal cell and liver malignancies. *American Journal of Medicine*, **51**, 767-771.
- Swets, J. A., Tanner, W. P., and Birdsall, J. G. (1964). In *Signal Detection and Recognition by Human Observers*, edited by J. A. Swets, p. 12. Wiley, New York.
- Teres, J., Bertran, A., and Mas, A. (1970). α_1 -Feto-protein in Spain (letter). *Lancet*, **2**, 465.
- Terry, R. (1952). Risks of needle biopsy of the liver. *British Medical Journal*, **1**, 1102-1105.
- Zamcheck, N., and Klausenstock, O. (1953). Liver biopsy. II. The risk of needle biopsy. *New England Journal of Medicine*, **249**, 1062-1069.
- Zawadzski, Z. A., and Kraj, M. A. (1974). Alpha-feto-protein in hepatocellular disease and neoplastic disorders. *American Journal of Gastroenterology*, **61**, 45-52.