

## 《講 座》

## 医用画像における臨床的有効度の評価

## ——(I) 決定マトリックスによる評価因子——

飯 沼 武\*

## I. はじめに

近年、医用画像をめぐる話題は非常に多い。総合画像診断だとか、クリニカル・エフィカシーだとか多くの言葉が文献にぎわしている。これは XCT の発明を機に医学、医療に導入された様々な方式の画像情報の洪水によるのではなかろうか。人体の内部構造を体外から計測する手段は、長い間 X 線写真だけに限られてきたが、1950 年ごろから RI イメージが加わり、1965 年ごろから超音波像、1972 年から XCT 像が参入するといった状況で、今も多様なエネルギー媒体を用いた人体内部の構造や機能の映像化が研究されており、それらが続々と加わってくることは明らかである。

XCT の出現によって RI イメージ診断は大きなインパクトを受け、検査件数も脳シンチグラフィを中心に減少した。そのこともあって、RI イメージ診断の臨床的有効度 (clinical efficacy) や費用・効果 (cost-effectiveness) などを XCT との関連で評価しようとの気運が急速に高まった。また、RI イメージ以外の他の画像診断法、特に通常の X 線診断や XCT 画像についても同様な評価を行なおうとする動きが始まっている。また、超音波像では未だ、定量的評価の問題は具体的には始まっていないが、技術的な進歩が成熟に達するにつれて、いずれこの問題を取り上げざるを得なくなるであろう。

実は、これら多くの医用画像に共通することは、これらの画像情報が病気の診断という情報に変換される際に医師による読影という過程を経由することである。もちろん、画像情報を直接、電子計

算機に読みこみ、自動的に病気を診断しようとする研究もあり、パターン認識 (pattern recognition) とよくいわれているが、これらについては未だ研究段階であり、実用的に多くの医用画像に使えるようになるのは当分先のことといわねばならない。

そこで、現時点では人間 (医師) を組みこんだ医用画像診断系の評価を定量的に行なうことがどうしても必要である。一般に医学における診断は不確実な情報のもとにおける意志決定であるため、わずかとはいえ、誤まりを避けることは不可能であり、医用画像の診断においても、例外ではない。また、読影診断のように人間が介在するシステムではその人の性格、能力、態度などによって誤まりの割合が異なり得るし、同一人でも日によって変動する。そこでこれらの割合を定量的に把握しておくことが医療の質の向上のために最も重要であるといっても過言ではない。

本講座では、今まであまり追求されていなかった読影診断の精度の問題を取り上げることにした。読影診断の精度を定量的に測る方法にいくつかの理論があるが、それを先に述べると読者諸兄がいやがさすと困るので、まず具体的に肝シンチグラムをある方式に従って読影したという想定のもとに、その結果をどのように整理して、診断精度の計算に進むかを説明し、そこで、決定マトリックスの作成法について述べることにする。具体的な話の後に、一般論としての誤診の考え方、決定マトリックス、ROC 曲線、ベイズの定理や最適な意志決定などに進むこととしたい。

最後にもう一度強調したいことは、本講座で述べる方法論は核医学イメージに限らず、X 線写真などの他の医用画像に適用可能であることである。

\* 放射線医学総合研究所臨床研究部

諸者諸兄におかれては、ここで述べる方法論を、多くの画像情報に応用するよう努力してもらいたい。

## II. 肝シンチグラムの読影

ここでは日本アイソトープ協会核医学開発委員会エフィカシー1小委員会(町田喜久雄委員長)で著者らも協力して行なった肝シンチグラムの読影実験の結果を具体例として利用する。

同委員会では、東京付近の8施設より確定診断の付いている肝シンチグラム症例を404例収集した。症例の収集の際は(a)正常例、局在性病変(SOL)例およびび慢性病変例が約1/3ずつ含まれること、(b)シンチグラムは最低、正面、後面、右側面が撮影してあるものに限定した。

収集された症例の内訳をTable 1, 2および3に示す。まず、Table 1では用いられた放射性医薬品の分布と、撮像装置の分布を示す。医薬品ではTc-フチン酸が圧倒的に多かったが、撮像装置ではカメラが404症例のうち約300例であった。Table 2は確定診断を付けた手技についての分布である。手術によるものが最も多く、次いで臨床診断、剖検の順であった。Space Occupying Lesion (SOL)の大きさが手術例と剖検例で可成り異なっており、SOLの検出において重大な問題となることがわかった。Table 3は確定診断名の頻度分布である。正常148例、SOL 118例、び慢性疾患124例、肝胆道疾患11例、そのほか3例で、ほぼ前に

Table 1 用いられた放射性医薬品と撮像装置

$^{99m}\text{Tc}$ phytate	378
$^{99m}\text{Tc}$ Sn colloid	19
Others	7
Scanner	95
Anger Camera	309
Scanner + Camera	2

Table 2 確定診断のための手技

Autopsy	92
Surgery	161
Endoscopy	45
Biopsy	77
Angiography	53
CT	31
Clinical diagnosis	102

Table 3 収集した全症例における肝疾患の分布

Normal	148
SOL	118 (total)
Hepatoma	18
Hepatoma with cirrhosis	18
Metastasis	69
Others	13
Bileduct abnormality	11
Diffuse live disease	124 (total)
Acute hepatitis	13
Chronic hepatitis	42
Liver cirrhosis	55
Others	14
Others	3
Total	404

### 6. 肝シンチグラム診断

(1) 異常 (以下有のときチェック)	有 <input type="checkbox"/> 1	無 <input type="checkbox"/> 0	47		
(2) 局在性病変	有 <input type="checkbox"/> 3	有の疑 <input type="checkbox"/> 2	無の疑 <input type="checkbox"/> 1	無 <input type="checkbox"/> 0	48
(4) 胆道系異常	有 <input type="checkbox"/> 2	有の疑 <input type="checkbox"/> 3	無の疑 <input type="checkbox"/> 1	無 <input type="checkbox"/> 0	50
(6) びまん性肝疾患 (肝硬変は除く)	有 <input type="checkbox"/> 3	有の疑 <input type="checkbox"/> 2	無の疑 <input type="checkbox"/> 1	無 <input type="checkbox"/> 0	52
(3) 肝硬変	有 <input type="checkbox"/> 3	有の疑 <input type="checkbox"/> 2	無の疑 <input type="checkbox"/> 1	無 <input type="checkbox"/> 0	49
(5) 肝外性病的圧迫	有 <input type="checkbox"/> 3	有の疑 <input type="checkbox"/> 2	無の疑 <input type="checkbox"/> 1	無 <input type="checkbox"/> 0	51
(7) その他の異常	有 <input type="checkbox"/> 3	有の疑 <input type="checkbox"/> 2	無の疑 <input type="checkbox"/> 1	無 <input type="checkbox"/> 0	53

Fig. 1 肝シンチグラム報告シートにおける肝シンチグラム診断の項目



述べた症例収集の条件を満足していた。

これらの症例を提供した8施設の11名の核医学専門医が50症例/月の速度で404例を読影した。読影は「肝シンチグラム報告シート」といわれる電子計算機入力用のワークシートに従って行なわれた。シートには患者番号、読影医師番号、肝所見として、① RI 集積、②変形および位置異常、③腫大、萎縮、④病的欠損、④生理的欠損など、脾所見として①描出の有無、②腫大、③欠損など、最後に肝シンチグラム診断の項目が含まれている。ここでは後の解析で使う肝シンチグラム診断の項目だけを Fig. 1 に示す。各項の内容は(1)全体としての異常有無、次いで(2)局在性病変から(7)その他の異常までが各論としてあり、(2)~(7)は「有」「有疑」「無疑」「無」の4段階に分けられている。これらは読影診断の際の確信度(confidence level)に当たり、後の ROC 解析に重要な役割りを演ずることになる。

読影は基本的には404例に対し、11名の医師によって行なわれたので、合計4444回の読影が行なわれたのであるが、後の解析では自施設の症例に対する読影結果を除いている。その理由は、自施設の症例については各医師とも最終結果を記憶しておられる可能性が大きいと判断したためである。従って、読影結果はすべて他施設の症例によるものであり、記憶によるバイアスは全くない。

また、ここで示す結果は肝シンチグラム診断における局在性病変(SOL)の検出に限ることとし、11名の医師の平均値と、ある特定の医師1名の結果を示すこととする。まず、Table 4に医師AによるSOLのシンチグラム診断と確定診断の比較を示す。医師Aは404症例のうち、364例のみ診断しているが、40症例は自施設の症例として除外したためである。

364症例のうち、確定診断の結果としてSOL有とされている症例が113例、SOL無(正常およびその他の疾患を含め)とされている症例が251例である。ここでは肝の部位、SOLの大きさ、撮影装置などを分類せずに全て一括した成績を示す。このような症例に対し、医師Aが「有」と診断

Table 4 局在性病変の肝シンチグラム診断

(医師 A)

シンチグラム 診断	確定診断	
	有	無
有	75	9
有の疑	15	33
無の疑	9	68
無	14	141
計	113	251
総計	251+113=364	

Table 5 局在性病変の肝シンチグラム診断

(11名の合計)

シンチグラム 診断	確定診断	
	有	無
有	876	148
有の疑	128	268
無の疑	46	312
無	190	2037
計	1240	2765
総計	1240+2765=4005	

した症例のうち、確定診断「有」のもの75例、「無」のもの9例が含まれる。続いて「有の疑」の診断では、確診「有」が15例、「無」が33例、診断「無の疑」では「有」9例、「無」68例、診断「無」では「有」14例、「無」141例が含まれていた。

一方、Table 5は11名の全医師の合計としてのSOLのシンチグラム診断と確定診断の比較である。各医師によって自施設の症例数が異なるため、11に割り切れる数にはなっていないが、全医師の診断した症例のうち、自分の症例を除いた者で、確定診断「有」が1,240例、「無」が2,765例、総計4,005例である。シンチグラム診断については医師Aの場合と同様に考えてよく、この成績は医師11名の平均的な値と言える。

### III. 正診率を表わす種々の因子

さて、本節では Table 4, 5 の読影診断と確定診断の比較結果を利用して、正診率を表わすいくつかの重要な因子を求めて見ることにしよう。

**Table 6** 医師 A の SOL 診断の決定マトリックス  
(表4より)

シンチグラム 診断	確定診断		計
	有 D+	無 D-	
「所見有」 有の疑} T+	75+15=90	9+33=42	132
「所見無」 無の疑} T-	9+14=23	68+141=209	232
計	113	251	364

**Table 7** 決定マトリックス (一般化した表現)

読影診断	確定診断		計
	有病 (D+)	無病 (D-)	
所見有 (T+)	a	c	a+c
所見無 (T-)	b	d	b+d
計	a+b	c+d	a+b+c+d

### 1. 決定マトリックス (decision matrix) を 利用する正診率の算出法

まず, Table 4 の結果に対し, シンチグラム診断において, 確信度の大きい「有」と「有の疑」を1グループとし, 小さい「無の疑」と「無」を1グループとして, 2つに分けてしまう. この分け方は, 「有」とそれ以外に分けてもよいし, 「有」から「無の疑」までを1グループに, 「無」を第2のグループとしてもよい. いずれにしても確信度によって2つに分割することを意味する.

そうすると, Table 4 は Table 6 のように4つのカテゴリーに分類されることになるであろう. すなわち, 確定診断「有」(D+と書く)の全症例113例のうち, シンチグラム診断「有」と「有の疑」(T+と書く)は90例, 他の23例はシンチグラム診断「無の疑」と「無」(T-と書く)とされたことを示している. それに対し, 確定診断「無」(D-)の全症例251例のうち, T+とされたもの42例, T-とされたもの209例である. この Table 6 は一般に病気を診断する場合の最も簡単な判別法である. そこでは診断す可き疾患(この場合, SOL)を「有」(D+)と「無」(D-)に分ける(読影実験

では確定診断が付いているから分けられるが, 通常の診療では追跡によってのみわかる). それに対して, 読影診断も確信度によって2つに分け, 「所見有 (T+)」と「所見無 (T-)」とする. この表を決定マトリックスと呼んでいる.

ここで話しを少し一般化しよう. Table 6 を一般化して表わした決定マトリックスを Table 7 に示す. そこでは確定診断を「有病 (D+)」と「無病 (D-)」の2つに分けている. ただし, ここでいう「有病 (D+)」とは今, 問題としている特定の疾患が有ることを言うのであって, 一般に病気が有ることを指すのではない. また「無病 (D-)」も病気が無く, 正常であるという意味ではなく, 今検出しようとしている病気が無いことを示すものであることに注意して頂きたい. Table 6 では「有病」は SOL の存在であり, 「無病」は SOL が無いことであって, この中には肝硬変などが含まれているのである. 一方, 読影診断の方も「所見有 (T+)」と「所見無 (T-)」の2種のみに分類する. ここでいう「所見無 (T-)」もある確信度以下の診断という意味であることに注意されたい. Table 4, 5 のように読影診断の確信度には通常複数個の段階があるので, どの段階で2つに分割するかによって, 4つのカテゴリーに入る症例数が異なってくるのがわかる.

さて, Table 7 からいくつかの正診率に関係する因子を求めることにする. まず, 対象とする症例は全部で  $a+b+c+d$  であり, おおの4つのカテゴリーに分けられている. ところが, おおのこのカテゴリーの意味する内容は大いに異なっている.

a は有病症例のうちで, 正しく「所見有 (T+)」と診断された症例で, 英語では true positive, 筆者らの提案した用語では「有病正診」<sup>1)</sup> 例という. Table 6 の例で述べれば, 確定診断「SOL 有」のうちで, 正しく「所見有」とされた症例であり,  $a=90$  例であることが容易にわかるであろう.

b は有病症例のうちで, 誤まって「所見無 (T-)」と診断された症例で, false negative, 「有病誤診」<sup>1)</sup> 例である. Table 6 では  $b=23$  例である.



cは無病症例のうちで、誤まって「所見無(T+)」と診断された症例で、false positive, 「無病誤診」<sup>1)</sup>例である。Table 6 では c=42 例である。

dは無病症例のうちで、正しく「所見無(T-)」と診断された症例で、true negative, 「無病正診」<sup>1)</sup>例である。Table 6 では d=209 例である。

Table 7 の4つのカテゴリーに分けられた症例数をもとに診断の質を表わす重要な因子がいくつか導かれる。以下に筆者らが提案した用語<sup>1)</sup>と英語を併記して示す。

(1) 有病正診率 (true positive rate: TPR):  $TPR = a/(a+b)$ . sensitivity ともいわれており、有病症例(D+)のうちで、所見有(T+)とされた者の割合、条件付き確率 (conditional probability) という形式で表わすと、 $P(T+|D+)$ 。

(2) 有病誤診率 (false negative rate: FNR):  $FNR = b/(a+b)$ . 有病症例(D+)のうちで、所見無(T-)とされた者の割合。  $P(T-|D+)$ 。(1)の式よりわかるように、 $FNR = 1.0 - TPR$ 。

(3) 無病誤診率 (false positive rate: FPR):  $FPR = c/(c+d)$ . 無病症例(D-)のうちで、所見有(T+)とされた者の割合。  $P(T+|D-)$ 。

(4) 無病正診率 (true negative rate: TNR):  $TNR = d/(c+d)$ . 無病症例(D-)のうちで、所見無(T-)とされた者の割合で、specificity ともいわれている。  $P(T-|D-)$ 。(3)の式より、 $TNR = 1.0 - FPR$ 。

(5) 尤度比 (likelihood ratio: L):  $L = TPR/FPR = a \cdot (c+d)/c \cdot (a+b)$ 。これは有病正診率と無病誤診率の比と定義されている。

(6) 正確度 (accuracy: A):  $A = (a+d)/(a+b+c+d)$ 。全症例のうちの正診された者の割合を表わす。

この結果、(1)~(4)で示された4つのカテゴリーのうち、(1)(2)のいずれかと、(3)(4)のいずれかの2つが決まれば、他は自動的に決まってしまうことがわかる。

これらの因子は診断医の能力や診断手技の質の良さを評価する方法を提供する。例えば、ある診断法はTPRが1.0に近く、FPRが0.0に近いほ

ど、その質がよいことを意味する。また、同一の症例群について読影を行なった医師の結果を比較したところ、医師Aが医師Bに比して、TPRが大きく(従って、FNRが小さく)、TNRも大きい(従って、FPRも小さい)場合には医師AはBに比して、診断能力が高い。

Table 6の例を使って、実際に(1)~(6)の因子を計算してみると、次のようになる。

$$TPR = \frac{90}{113} = 80\%$$

$$FNR = \frac{23}{113} = 20\%$$

$$FPR = \frac{42}{251} = 17\%$$

$$TNR = \frac{209}{251} = 83\%$$

$$L = \frac{0.8}{0.17} = 4.7$$

$$A = \frac{90+209}{364} = 82\%$$

ここで、Table 5の結果を、Table 6と同様に決定マトリックスに組み変え、(1)~(6)の因子を計算すると、SOL検出に対する11名の医師の平均的な値が求められることになる。

読者諸兄の練習問題としてやってもらいたい、結果だけ以下に示す。

$$TPR = 81\% \quad FNR = 19\%$$

$$FPR = 15\% \quad TNR = 85\%$$

$$L = 5.4 \quad A = 84\%$$

このように分類した決定マトリックスによって正診率をみると、医師A (Table 6)は全医師の平均値よりも診断能力がやや低いことがわかる。

## 2. Predictive value とベイズの定理

本節では、Table 7の決定マトリックスを別の見方より評価してみることにする。実はIII-1節では確定診断における有病症例と無病症例に対する割合という形で、様々な正診率に関係する因子を定義したが、ここでは読影診断の結果側から見た重要な因子を求めてみよう。たとえば、読影診断で「所見有(T+)」とされた症例はTable 7ら

a+c であるが, そのうち正しく病気があった者は a のみであるという場合などをいう. 読影診断の側から見た因子にも次の4種があり, その名称には筆者らの提案したものを使用して頂くことにする<sup>1)</sup>.

1) 有徴正診率: これは所見有 (T+) と診断された者のうちで, 正しく有病 (D+) であった者の割合を示し, 条件付き確率で表わせば,  $P(D+|T+)$  と書ける. この値は次のように計算される.

$$P(D+|T+) = a/(a+c)$$

英語では predictive value of positive test results と呼ばれている.

2) 無徴正診率: 所見無 (T-) と診断された者のうちで, 正しく無病 (D-) であった者の割合を示し,  $(P(D-|T-))$ , 次のように計算される.

$$P(D-|T-) = d/(b+d)$$

英語では predictive value of negative test results と呼ばれている.

3) 有徴誤診率: 所見有 (T+) と診断された者のうちで, 誤まって無病 (D-) であった者の割合を示し,  $(P(D-|T+))$ , 次のように計算される.

$$P(D-|T+) = c/(a+c)$$

(1) で示した式より, 次の関係のあることがわかる.

$$P(D-|T+) = 1.0 - P(D+|T+)$$

4) 無徴誤診率: 所見無 (T-) と診断された者のうちで, 誤まって有病 (D+) であった者の割合を示し,  $(P(D+|T-))$ , 次のように計算される.

$$P(D+|T-) = b/(b+d)$$

(2) で示した式より, 次の関係のあることがわかる.

$$P(D+|T-) = 1.0 - P(D-|T-)$$

ここで述べた4種の因子は読影診断を行なった結果のうちで, どのぐらい正診されたかを定める因子であって, 診断に際しては III-1 で述べた諸因子よりも直接的な値であると言えよう. III-1 の諸因子との関係については後で詳しく述べる.

次に, 医師 A の結果 (Table 6) を用いて, 上記 (1)~(4) の諸因子を算出してみよう.

$$\text{有徴正診率: } P(D+|T+) = \frac{90}{90+42} = 68\%$$

$$\text{無徴正診率: } P(D-|T-) = \frac{209}{23+209} = 90\%$$

$$\text{有徴誤診率: } P(D-|T+) = \frac{42}{90+42} = 32\%$$

$$\text{無徴誤診率: } P(D+|T-) = \frac{23}{23+209} = 10\%$$

また, III-1 と同じように, Table 5 についても計算でき, 結果は次のようになる. (読者の検算をお願いしたい.)

$$P(D+|T+) = \frac{1004}{1420} = 72\%$$

$$P(D-|T-) = \frac{2349}{2585} = 91\%$$

$$P(D-|T+) = 28\%$$

$$P(D+|T-) = 9\%$$

これらの因子の意味するところは, 例えば, 有徴正診率68%とは医師 A が「所見有 (T+) とした症例のうち, 68%が正しく問題とする病気をもっていた」ということである.

ここで一般的に有徴正診率などの因子と III-1 の有病正診率などとの関係を表わすことを考えてみよう. この両者の関係を知ることは正診率の2つの側面を把握する上で重要だからである. 実はその関係を示す式がベイズの定理と言われており, 次のように表わされる.

$$P(D+|T+) = \frac{TPR \times P(D+)}{TPR \times P(D+) + FPR \times P(D-)}$$

確率で表わすと,

$$= \frac{P(T+|D+) \times P(D+)}{(P(T+|D+) \times P(D+)) + (P(T+|D-) \times P(D-))} \quad (1)$$

$$P(D-|T-) = \frac{TNR \times P(D-)}{TNR \times P(D-) + FNR \times P(D+)}$$

確率で表わすと,

$$= \frac{P(T-|D-) \times P(D-)}{(P(T-|D-) \times P(D-)) + (P(T-|D+) \times P(D+))} \quad (2)$$

ここで,  $P(D+)$  は今, 問題としている患者全体のうちの有病者 (D+) の割合をいい,  $P(D-)$  は逆に無病者 (D-) の割合をいう. 従って,



$P(D-)=1.0-P(D+)$  であり、これらの因子を事前確率 (prior probability) と称する。それに対し、算出された条件付き確率  $P(D+|T-)$  などを事後確率 (posterior probability) と称する。また、 $P(D+)$  は患者全体中の有病者の割合だから、有病率 (prevalence) とも言われている。

ベイズの定理がなぜ成り立つかを上の2式に従って考えて見よう。(1)式の分母を見ると、有病者の割合 ( $P(D+)$ ) に、有病正診率 (TPR) を乗じ、次に無病者の割合 ( $P(D-)$ ) に無病誤診率 (FPR) を乗じ、両者の和をとっている。ここで前者は有病者のうちで、正しく所見有 ( $T+$ ) と診断されたものであり、後者は無病者のうちで、誤まって所見有 ( $T+$ ) と診断されたものであるから、この両者の和は所見有 ( $T+$ ) と診断された者の合計である。一方、分子は分母の第1項に対応し、有病者のうちの所見有のものであるので、(1)式はまさに有徴正診率そのものであることがわかるであろう。同様に考えて見ると、(2)式も所見無 ( $T-$ ) と診断された者全員 (分母) のうちの、無病者 ( $D-$ ) で所見無とされた者 (分子) であるから、無徴正診率そのものである。

(1)式は、事前確率  $P(D+)$  の有病率であったものが、ある検査を行なった結果、もし所見有 ( $T+$ ) だとしたら、有病の確率は  $P(D+|T+)$  という事後確率に変化することを示している。 $P(D+|T+)$ ,  $P(D-|T-)$  とも1.0に近いほど、その診断システム (人間も含め) の性能は秀れているわけである。

(1), (2)式で注意すべきことは、事前確率 (有病率)  $P(D+)$  が入ってくることである。従って、TPR と FPR が一定であったとしても、事前確率の値によって、有徴正診率などの事後確率が変わることである。

再び、医師Aの結果 (Table 6) を用いて、(1), (2)式を算出して見よう。まず、事前確率  $P(D+)$  を求める必要があるが、この場合は医師Aの見た症例が全例であるとすれば、全症例中の有病者の割合であるから、次のように計算される。

$$P(D+)=\frac{113}{364}=0.31$$

$$P(D-)=1.0-P(D+)=0.69$$

医師AについてのTPRなどは前に求められているので、それらを代入すると、医師Aの有徴正診率と無徴正診率は次のように計算される。

$$P(D+|T+)=\frac{0.8 \times 0.31}{0.8 \times 0.31 + 0.17 \times 0.69}=0.68$$

$$P(D-|T-)=\frac{0.83 \times 0.69}{0.83 \times 0.69 + 0.2 \times 0.31}=0.90$$

当然のことであるが、前に計算した結果と一致している。Table 5についても、 $P(D+)$  を求め、(1), (2)式により計算することを練習問題としてやってもらいたい。

ここで、医師AについてTPRとFPRが一定である (従ってFNRとTNRが一定) として、 $P(D+)$  が変化する場合を考えてみよう。例えば、医師Aが診療している患者層が変化した場合や、他の病院に転勤して、そこの患者が前のところとは異なっている場合には、有病率  $P(D+)$  の変化がおこることが十分予想される。その際、医師Aの読影診断能力としてのTPR, FPRが変わらなかったとしても、有徴正診率などの事後確率は変わるのである。

もし、患者のうちの有病者の割合がTable 6の場合より低く、 $P(D+)=0.1$  (従って  $P(D-)=0.9$ ) であるとしよう。その場合、医師Aの事後確率は次のように変わる。

$$P(D+|T+)=\frac{0.8 \times 0.1}{0.8 \times 0.1 + 0.17 \times 0.9}=0.34$$

$$P(D-|T-)=\frac{0.83 \times 0.9}{0.83 \times 0.9 + 0.2 \times 0.1}=0.97$$

$P(D+)$  が0.001 (1,000人に1人) から0.5 (2人に1人) に変化した時に、有徴正診率と無徴正診率がどのように変動するかをTable 8に示す。ここではTPR=0.99, FPR=0.01というかなり理想に近い診断システムの場合とTable 6で示した医師Aの場合について計算した。理想的な診断系でさえ、 $P(D+)$  が集検の場合のように1,000人に1人ぐらいに減少すると、有徴正診率は10人に1人、

**Table 8** 有病率 ( $P(D+)$ ) の変化による有徴正診率と無徴正診率の変動 (理想的な診断システムの場合と医師 A の場合)

$P(D+)$	理想的に近い診断システム TPR=0.99 FPR=0.01		医師 A の場合 TPF=0.80 FPR=0.17	
	有徴正診率	無徴正診率	有徴正診率	無徴正診率
0.001	0.09	1.0	0.005	1.0
0.01	0.5	1.0	0.045	1.0
0.02	0.67	1.0	0.088	1.0
0.05	0.84	1.0	0.11	0.999
0.1	0.92	0.999	0.34	0.97
0.2	0.96	0.997	0.54	0.94
0.5	0.99	0.990	0.82	0.81

すなわち、「所見有 (T+)」の中に、有病者は10人のうち1人しかいない。9人は無病者である。 $P(D+)$  が小さい場合は有徴正診率に大きい影響を及ぼすことがわかる。Table 8 の結果は (1), (2) 式を使って、ご自分で計算してみることをおすすめしたい。いずれにしてもご自分の診療しておられる患者層のマスとしての有病率を概略でもよいから把握していることはきわめて大切なことなのである。

ベイズの定理は (1), (2) 式よりもさらに一般化でき、2種類以上の相互に独立な症状、所見などから複数の疾患を鑑別することにも使えるが、詳細はよい参考書<sup>2)</sup>があるので、それを見られたい。また、最近の日本医放会誌にベイズの定理を応用した興味ある論文がのっている<sup>3,4)</sup>。

### 第 I 部のまとめ

本稿では肝シンチグラムの読影診断の具体例と確定診断との比較から入り、決定マトリックスの作成法、各種の正診率の求め方を説明した。特にわれわれの提案した用語を中心にその概念をはっきりさせ、肝シンチグラム診断の例を使って数値計算を行なった。外国の文献でも有病正診率のグループと有徴正診率のグループの間の混乱が見受けられるので、読者諸兄におかれても、その辺を

十分につかんでももらいたい。ベイズの定理については最も簡単な  $2 \times 2$  決定マトリックスの場合のみ説明したので、興味ある読者はさらに勉強していただきたい。

なお、肝シンチグラム読影実験の結果については、今後、多くの観点から分析を続ける予定であり、本稿に示したものはその一例に過ぎない。この種の実験の評価に当たっては、もとの症例の内容によって結果が全く変わり得ることを注意しておこう。われわれとしては、その点につき実験計画の始めに相当に配慮したつもりである。

参考文献は、本講座の最終稿にまとめてあげることにするが、第 I 部で引用した文献については下記に示す。次回は ROC 解析から、最適意志決定、および臨床の有効度の測定法などにつき述べることとする。

### 文 献

- 1) 舘野之男, 飯沼 武: 誤診などに関する用語の提案. 医学のあゆみ 97: 424-426, 1976
- 2) 斎藤嘉博: 決定のはなし. 日科技連出版, 東京, 1969
- 3) 小川史顕: 胸部X線診断情報処理システムに関する研究. 日本医放会誌 39: 862-877, 1979
- 4) 小牧専一郎: 原発性脳腫瘍の CT パターン解析およびそれによる自動診断の可能性の検討. 日本医放会誌 39: 1055-1069, 1979