DIAGNOSIS, AND BAYES' THEOREM

Sir,—It was encouraging to see three contributions to the clinical application of Bayes’ theorem in The Lancet of Feb 9 (pp 326, 329, and 348). It is certainly much easier to use the odds ratios rather than the probabilities and easier still to use the nomogram of Fagan. I have used a copy of this nomogram pinned above my desk since it was published in 1975.

However, it is time that we abandoned the use of fixed levels of sensitivity and specificity. The notion that a test result within the normal reference range always indicates normality and that a result outside it indicates disease is usually fallacious. Borderline results suggest the possibility of false-positive or false-negative interpretations. There is no need to use a single cut-off point between normality and disease. Rare diseases which are dangerous to treat require stringent acceptance criteria. Common diseases which are safe to treat allow a larger false-positive error rate. There is no need to use a single cut-off point outside it indicates disease is usually fallacious. Borderline results suggest the possibility of false-positive or false-negative interpretations. There is no need to use a single cut-off point between normality and disease. Rare diseases which are dangerous to treat require stringent acceptance criteria. Common diseases which are safe to treat allow a larger false-positive error rate. There is always an inverse relation between false-positive and false-negative errors. The optimum cut-off point may be determined by plotting true-positive fractions against false-positive fractions for selected cut-off points of the test. For any point on the ROC curve, the ratio of these fractions (corresponding to the probability of the test being positive in the presence of disease divided by the probability of the test being positive (in the absence of disease) is the likelihood ratio of a positive test result. Prior odds result from multiplying this by the prior odds, a quick way of solving the Bayes’ formula.

Clinicians and pathologists should cooperate in obtaining the necessary data. “Disease” cases must of course be identified by methods different from the test under consideration.

It is also possible to apply ROC methods to clinical trials. The analogy of false-positive (type I) and false-negative (type II) errors in diagnosis to those in therapeutic trials is obvious. Instead of using a fixed a value (usually 0·05) and collecting enough numbers to achieve a certain power (1−β), the results may be inspected at any point and the likelihood ratio of the difference between the treatments (1−β/a) calculated. Then small differences in important indices such as mortality would not be overlooked.

It is a great pity that clinicians seem more reluctant to use the potentially greater contribution to diagnosis of simple arithmetic than the expensive and often misleading results achieved by high technology. Perhaps this reluctance will persist until request forms can be despatched to a “numbers laboratory”, as in Boston. It would be preferable, however, for the manifold institutions of undergraduate and postgraduate education up and down the country to provide courses in medical decision making. These would be more enjoyable and more cost-effective than seminars dealing with the latest (but not the last) management reorganisation.

Royal Devon & Exeter Hospital (Wonford),
Exeter EX2 5DW
G. H. HALL


NUTRITIONAL STATUS AND INFECTIONS IN ETHIOPIAN IMMIGRANTS

Boys who were younger than 5 years and girls who were younger than 6 years were selected as an initial sample of 2,500 children. These children were followed up for five years, and the data analysis was only done for the first two years. The anthropometry and physical findings were recorded at each follow-up. The anthropometric variables included body-weight as % of ideal, triceps skin fold in mm, total protein and albumin in g/dl.

SIR,-The Lancet of Feb 9 contains two articles (by Dr Balla and colleagues and by Dr Simel) on how to update uncertainty in the light of evidence. The reasoning of the first article is difficult to follow, but by using Simel’s techniques perhaps the argument can be clarified.

A patient has a disease D and then a symptom S occurs, which may be due to a complication C of the disease or to a new condition N. A diagnostic test for C is negative (T−) and we want to know whether a complication rather than a new condition has occurred. Our uncertainty may be expressed as the odds on C relative to N, given S and T−. If the symptom and test result are independent features of the underlying cause Simel’s approach shows that the odds on C relative to N is (likelihood ratio of T−) × (likelihood ratio of S) (prior odds on C relative to N), which may be written as:

$$P(T−|C)P(S|C)P(C)$$
$$P(T−|N)P(S|N)P(N)$$

Using the estimates provided by Balla et al we calculate the final odds to be

<table>
<thead>
<tr>
<th>Patient (age, sex)</th>
<th>Weight</th>
<th>Triceps</th>
<th>Total protein</th>
<th>Albumin</th>
<th>Diagnoses†</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (45, F)</td>
<td>66%</td>
<td>4.0</td>
<td>5.8</td>
<td>2.3</td>
<td>TB, schistosomiasis</td>
</tr>
<tr>
<td>2 (29, F)</td>
<td>86%</td>
<td>11.0</td>
<td>7.3</td>
<td>3.6</td>
<td>TB, syphilis, latent</td>
</tr>
<tr>
<td>3 (22, M)</td>
<td>72%</td>
<td>4.6</td>
<td>6.2</td>
<td>2.2</td>
<td>TB, syphilis, latent; hookworm</td>
</tr>
<tr>
<td>4 (30, M)</td>
<td>85%</td>
<td>7.5</td>
<td>5.2</td>
<td>2.3</td>
<td>Typhoid</td>
</tr>
<tr>
<td>5 (39, M)</td>
<td>74%</td>
<td>4.0</td>
<td>4.7</td>
<td>2.3</td>
<td>Typhoid</td>
</tr>
<tr>
<td>6 (58, M)</td>
<td>56%</td>
<td>3.0</td>
<td>5.0</td>
<td>1.9</td>
<td>Paratyphoid B; shigellosis; hookworm</td>
</tr>
<tr>
<td>7 (40, F)</td>
<td>61%</td>
<td>3.0</td>
<td>4.7</td>
<td>2.0</td>
<td>Paratyphoid C</td>
</tr>
<tr>
<td>8 (20, F)</td>
<td>92%</td>
<td>12.0</td>
<td>7.4</td>
<td>3.1</td>
<td>Malaria; ascariasis</td>
</tr>
<tr>
<td>9 (18, F)</td>
<td>84%</td>
<td>8.5</td>
<td>6.6</td>
<td>3.5</td>
<td>Malaria; schistosomiasis; pneumonia</td>
</tr>
<tr>
<td>10 (40, M)</td>
<td>80%</td>
<td>4.5</td>
<td>7.1</td>
<td>3.4</td>
<td>Staphylococcal abscess</td>
</tr>
<tr>
<td>11 (20, F)</td>
<td>87%</td>
<td>7.0</td>
<td>6.4</td>
<td>2.5</td>
<td>Bronchopneumonia</td>
</tr>
<tr>
<td>12 (38, F)</td>
<td>99%</td>
<td>10.0</td>
<td>6.7</td>
<td>3.7</td>
<td>Pneumonia</td>
</tr>
<tr>
<td>13 (25, F)</td>
<td>75%</td>
<td>5.0</td>
<td>6.5</td>
<td>5.2</td>
<td></td>
</tr>
</tbody>
</table>

*Body-weight as % of ideal; triceps skin fold in mm; total protein and albumin in g/dl.

†TB = tuberculosis; P = pneumonia; C = complications; N = new condition.
At no time up to now have we assumed that C and N cannot both be present or that one of them must be present. If we do make those assumptions then odds of “75 to 1” on can be changed to a probability for C of 75/(75+1)= 0-987. However, it seems much more intuitive to work entirely in odds. Balla et al found that the clinicians tended to forget about the initial odds when confronted with two pieces of negative evidence. This neglect of “base-rates” is well known in psychological work in probability assessment.

Balla and colleagues’ first formula:

\[
P(S) = P(S|N)P(N) + P(S|C)P(C)
\]

is appropriate only if it is assumed that one and only one of C or N must be present; this is not explicitly assumed until later in the article. The second formula, \(P(S|C)P(S|N)\), is incorrect, since the paragraph is concerned with the joint occurrence of C and N which, assuming independence, has probability \(P(C)P(N)\). Finally, the top two labels on the branches of the probability tree should be labelled “C and not N” and “N and not C”.

MRC Biostatistics Unit, Cambridge CB2 2QH

David J. Spiegelhalter


**This letter has been shown to Dr Balla and colleagues, whose reply follows. In Dr Møller-Petersen’s letter (Feb 9, p 348) the second term in the formula for efficiency should be \((P\times\text{specificity})\) and the legend to the nomogram should end “Similarly for P…”—E. L.

Sir,—We thank Dr Spiegelhalter for converting our probability estimates to odds. Whilst there are many who find working in probabilistic terms easy, others intuitively prefer to think in terms of odds. We do not believe that thinking in terms of odds is intuitively easier for most clinicians. It is indeed fortunate that two papers in the same issue of The Lancet should present two different approaches to the same topic. It should encourage clinicians to follow whichever mode of calculations they prefer, at the same time stressing that the results of such calculations will often differ from clinical intuition.

We believe that our second formula is appropriate because it provides an estimate of the probability that the symptom was produced jointly by the complication and the new disease. We did not attempt to estimate the joint occurrence of the two conditions as Spiegelhalter indicates. We believe that the extra notation suggested for the first two branches of our tree is unnecessary, since it is clear from the context that “C and not N” and “N and not C” are implied. Branches three and four deal with the situations where C and N are either present in combination or are both absent.

We thank Spiegelhalter for calling attention to the relation between our work and base-rate neglect. Our results demonstrate one more clinical context in which this psychological phenomenon can be elicited, showing that it is not confined to laboratory settings or artificial questions.

Neurology Unit, Prince Henry’s Hospital, Melbourne 3004, Australia; and Center for Educational Development, University of Illinois at Chicago, Chicago, Illinois, USA

John I. Balla
Robert Ianser
Arthur Elstein

SHEFFIELD SIDS DEATH RISK SCORE APPLIED TO AN IRISH POPULATION

Sir,—One approach to the prevention of sudden-infant-death syndrome (SIDS) is to attempt to identify infants at risk. This approach seems to have reduced the incidence of SIDS in Sheffield, UK, where the epidemiological characteristics associated with SIDS were used to construct a statistically weighted score which, when applied at birth and 4 weeks of age, identified 73% of subsequent SIDS victims. This encouraging report from Sheffield led us to evaluate this scoring system in all liveborn infants delivered in the Rotunda Hospital, Dublin, from Jan 1, 1979, to Dec 31, 1981.

Subsequent deaths from SIDS were identified from data supplied by the Medico-Social Research Board. As controls we selected the three infants born before and the two born after the SIDS infant. We followed the Sheffield scoring system except that part 2 was applied at 6 weeks because this is when a well-baby examination is done at this hospital. Home conditions were assessed by interview and/or a home visit. A diagnosis of SIDS was accepted where no adequate cause of death was found postmortem.

Of the 18 801 infants born alive 48 subsequently died from the SIDS (2.5% per 1000). The Sheffield scoring system had some predictive value but it was neither very sensitive nor very specific. At birth a score of over 500 identified 29% (14/48) of future SIDS victims and labelled 15% (29/192) of controls as at risk. At 6 weeks (scores above 745) these figures were 37-5% (18/48) and 14-5% (28/192), respectively. At best almost two-thirds of future SIDS victims are missed (false negatives) and 99-3% of infants identified as being at high risk of dying did not (false positives).

The attractiveness of a scoring system in early infancy to predict SIDS has led many investigators to attempt to construct their own scores, usually without success. The Sheffield score contributed to a sharp fall in the incidence of SIDS when used in conjunction with regular home visiting of "high risk" infants and was claimed to be cost effective. However, our data do not justify application of the Sheffield scoring system to an Irish population. Many of the factors found to be associated with SIDS in Sheffield when taken alone showed no positive correlation with SIDS in Dublin: in particular, we found no evidence that socioeconomic disadvantage or lack of breast feeding were significant factors. A modified score might be more suitable.

Department of Paediatrics, Rotunda Hospital, Dublin 1, Ireland

S. J. O’Brien
T. G. Matthews

CORRELATION BETWEEN FETAL DEFECTS AND EARLY GROWTH DELAY OBSERVED BY ULTRASOUND

Sir,—Pedersen and Mølsted-Pedersen have suggested that early fetal growth delay might be an index for high risk of serious congenital malformations in babies born to diabetic women. They call for a search for a similar association in non-diabetic pregnancies.

The three groups of women we studied met the following criteria: regular menstrual cycles of 28–30 days, no use of oral contraceptives for 3 months before pregnancy, singleton pregnancy, and measurement in our department by ultrasound of crown-rump length (CRL) at least once during the 7th to 14th weeks of pregnancy. 186 non-diabetic women who gave birth to normal malformed children during the two years 1982–83 met these criteria (controls) as did 72 insulin-dependent diabetic mothers (1976–83) and 16 non-diabetic mothers (1979–83) in whose babies major congenital malformations, chromosomal abnormalities, or inborn errors of metabolism were subsequently found.

CRL was usually measured on a Toshiba real-time linear array ultrasound machine (SAL 20A from 1978 to 1982, SAL 50A in 1983). Fetal age was estimated with Robinson and Fleming’s charts, and the results were expressed as the differences between ages as assessed by ultrasound and by menstrual dates.

Early growth delay was observed significantly more often in fetuses of diabetic mothers and fetuses with defects (table I). Of the