Pergamon

European Journal of Cancer Vol. 32A, No. 12, pp. 2042–2048, 1996

Printed in Great Britain
0959-8049/96 $15.00 + 0.00

PII: S0959-8049(96)00282-1

Review

The Diagnostic Information of Tests for the Detection of Cancer: the Usefulness of the Likelihood Ratio Concept

R.W.M. Giard[1] and J. Hermans[2]

[1]Department of Clinical Pathology, St Clara Hospital, Olympiaweg 350, 3078 HT Rotterdam and Section for Clinical Decision Analysis, University of Leiden, Leiden; and [2]Department of Medical Statistics, Faculty of Medicine, University of Leiden, Leiden, The Netherlands

No single test is perfect and without false-negative and/or false-positive results. Consequently, the clinician is perpetually confronted with incertitude about the true disease state of the patient. In oncology, these diagnostic errors may have harmful consequences for the patient. It is, therefore, imperative that the clinician knows how often these errors occur, which implies a quantitative evaluation of a test. With this knowledge, the test result must subsequently be interpretated within the clinical framework. Bayes' theorem provides a simple and useful mathematical model for the integration of measures of test performance and clinical data. Traditionally, sensitivity and specificity are used to describe test performance. However, this approach requires that the conclusion of the test is dichotomised into 'normal' and 'abnormal'. Few tests have a natural binary outcome. A test parameter that is applicable to all types of test outcome scales and, at the same time, provides the opportunity to determine the gain in diagnostic information by applying Bayes' theorem, is therefore mandatory. The likelihood ratio meets these conditions. The application of this concept for both the evaluation and the interpretation of various types of tests used in cancer patients is demonstrated. **Copyright © 1996 Elsevier Science Ltd**

Key words: likelihood ratio, test evaluation, test interpretation, sensitivity, specificity, Bayes' theorem

Eur J Cancer, Vol. 32A, No. 12, pp. 2042–2048, 1996

INTRODUCTION

IN ONCOLOGY, diagnostic errors may have great emotional impact. A false-negative test result gives unfounded reassurance and will cause an unnecessary delay of treatment. A false-positive test outcome means unnecessary severe distress, more testing and often unneeded—usually aggressive—therapy. The solution to this problem seems simple: the diagnostic tests applied should be brought to the level of perfection. However, the ideal diagnostic test, that unambiguously separates healthy from diseased, simply does not exist [1]. Consequently, in the diagnostic process a clinician is always confronted with uncertainty about the true state of the patient.

Correspondence to R.W.M. Giard.
Received 12 Apr. 1996; revised 19 Jul. 1996; accepted 25 Jul. 1996.

By its very nature, the diagnostic strategy is oriented at diminishing uncertainty. The clinician stepwise requests tests, hoping the outcome will enable him to confirm or to exclude a disease with more certainty than before. This process of using and reading diagnostic information can conveniently be modelled and formalised using some application of probability theory [2]. Traditionally, the concepts of sensitivity and specificity, to characterise test performance quantitatively, have been widely instituted in the medical literature. The estimation of these test parameters requires the systematic collection and analysis of test outcome data. By these indices we are informed about the probability of a test result given the disease state of a patient.

Although informative, the use of sensitivity and specificity has several practical drawbacks [3]. The likelihood ratio (LR) is an attractive addition to these conventional test characteristics. In this article, the concept and the appli-

cation of the likelihood ratio tailored to the diagnosis of neoplastic disease and its sequelae will be advanced.

TEST CHARACTERISATION: SOME LIMITATIONS OF SENSITIVITY AND SPECIFICITY

In oncology, tests are used for various reasons: to screen for cancer in asymptomatic individuals, to diagnose or to exclude cancer in symptomatic individuals, to detect metastatic spread, to monitor for relapse during follow-up, to direct therapy, or to ascertain complications of cancer treatment. (The word 'test' is used here in its broadest perception.) The leading diagnostic information is obtained from questioning and physical examination of the patient. This is followed by additional tests, most often haematology, clinical chemistry, imaging techniques and cyto- and histopathology. The upshot from testing is invariably expressed on some scale [4].

Occasional tests have a natural binary outcome, e.g. the presence or absence of a finding. Many tests however have a different upshot: their result is given on a continuous scale (e.g. concentration of a substance in the blood or serum) or as ordinal categories where a hierarchical sequence is noted (e.g. normal, suspicious, highly suspicious, definitively abnormal). Sometimes a test outcome itself is 'indeterminate' or 'uninterpretable'.

Suppose a patient has an enlarged cervical lymph gland with constitutional symptoms indicative of Hodgkin's disease. In the first exploratory diagnostic phase, an erythrocyte sedimentation rate (ESR) is determined. The outcome is 42 mm/h, which is abnormal. What does this outcome tell the clinician about the likelihood for this patient of really having Hodgkin's disease? Is Hodgkin's disease more likely then before, after this finding of ESR = 42 mm/h? To look formally to the diagnostic information of this test outcome for this patient, it is desirable to have information about the ESR distribution in 'proven' Hodgkin's patients and in patients who turned out (after a first suspicion) not to suffer from Hodgkin's disease. Table 1 gives (hypothetical) data for such information.

Using only (or mainly) the sensitivity and the specificity of an elevated ESR in this context has two disadvantages. First, for the calculation of sensitivity or specificity, the test outcome always has to be dichotomised into 'normal' and 'abnormal'. The choice of a cut-off point is arbitrary. In the given example, one might choose ESR $\leqslant$ 30 mm/h for 'normal' and ESR > 30 mm/h for 'abnormal', leading to a sensitivity of 40/160 = 0.25 and a specificity of 180/200 = 0.90. However, taking a lower cut-off value (e.g. 20 mm/h) leads to a higher sensitivity of 100/160 = 0.63 and a lower specificity of 150/200 = 0.75.

Table 1. Results from the determination of the ESR in patients with or without Hodgkin's disease (HD)

ESR (mm/h):	HD present *n* (%)	HD absent *n* (%)
0–10	30 (19)	90 (45)
11–20	30 (19)	60 (30)
21–30	60 (38)	30 (15)
31–40	20 (13)	10 (5)
41–50	10 (6)	8 (4)
51–60	5 (3)	2 (1)
>60	5 (3)	0 (0)
Total	160 (100)	200 (100)

When the conclusion from testing is not dichotomous, as is the case with most tests commonly used, the range of outcomes must be reduced to the two classes (normal, abnormal), which means loss of information due to the aggregating outcome categories or even the discard of non-positive and non-negative results. Table 1 shows that an outcome of ESR = 65 mm/h is more indicative of Hodgkin's disease than ESR = 42 mm/h.

Second and more importantly, the sensitivity and the specificity of a particular test are established in retrospect in a group of patients. The point of departure for the determination of these test characteristics is the true disease state (in our example: the presence or absence of Hodgkin's disease), which serves as a bench-mark for the test result. In the real-world situation of seeking the diagnosis, the physician does not know the true disease state of a patient prior to testing. Here, the starting point is the test result (e.g. ESR = 42 mm/h) and the doctor wants to know how the odds in favour or against disease may change with this particular test outcome. The past experience, summarised in Table 1, tells the clinician that ESR values between 41 and 50 occurred in 6 of the diseased cases and in 4 of the non-diseased cases. So the outcome 42 mm/h in this case favours disease by a ratio of 6/4 = 1.5. This ratio is more informative when seeking a diagnosis than is knowing either sensitivity, sensitivity or both. In combination with the prior probability on Hodgkin's disease, it will tell us how likely is the disease given the outcome ESR = 42 mm/h.

What is needed is a test characteristic applicable to the full range of results from the test applied, regardless of the output scale and that simultaneously relates the particular result to the distributions of the test outcomes in both the diseased population and the non-diseased population. The concept of the likelihood ratio (LR) fulfills these requirements and circumvents the limitations of sensitivity and specificity mentioned above.

Not only is the LR informative on a single test outcome, the LR is as appropriate for the interpretation of combinations of test results. In that situation, the same question with respect to the diagnostic information holds: given a particular combination of test results from a patient, how likely is the result given disease compared with non-disease? The LR of a combination of two (or more) independent tests is simply the product of the individual LR's. In case of dependent tests, other methods are available to come to a LR and so to a diagnostic appraisal of the test findings [5].

THE CONCEPT AND DEFINITION OF THE LIKELIHOOD RATIO (LR)

The concept of the LR is simple and revealing. The ratio expesses the probability of a particular test outcome in diseased patients divided by the probability of that outcome in non-diseased subjects. In other words: the LR is the ratio of two conditional probabilities (or relative frequencies) telling us how frequently this outcome has been seen in diseased patients compared with non-diseased patients. In our example: an ESR between 41 and 50 mm/h was 6/4 = 1.5 times more likely in patients with Hodgkin's disease compared with patients without it. It is akin to the risk ratio

used in epidemiology. The LR of a defined diagnostic outcome category **u** is therefore:

$$LR(u) = \frac{\text{probability of } u \text{ given presence of disease}}{\text{probability of } u \text{ given absence of disease}}$$

or in a more formal notation:

$$LR(u) = \frac{P[u|D]}{P[u|\text{non-D}]}$$

where $P[u|D]$ is the probability of **u** given the presence of disease D and $P[u|\text{non-D}]$ the probability of **u** given the absence of disease.

Since the numerator and the denominator are probabilities ranging from 0 to 1, the quantity $LR(u)$ will always be positive and takes values from 0 to infinity. From the general LR-formula, it can be seen that a test outcome with LR = 1 is virtually non-discriminating: the likelihood of this finding is equal in both the diseased and the non-diseased individuals. An LR well below 1 indicates that the possibility of the disease under investigation is less likely after the test result was obtained (ruling out of disease) while for LR > 1 the opposite is true (ruling in). The value of the LR, as can be seen from the formula, is unaffected by the prior probability (prevalence) of the disease.

LRs are always derived from examining samples of patients. A different sample will lead to a different LR-value. Thus, they are exposed to random variation, and the use of confidence intervals (CI) to allow for a critical evaluation of the estimates is, therefore, mandatory. The calculation of CI for sensitivity and specificity is straightforward. For LRs, several techniques are propagated. They are discussed in more detail elsewhere [6].

Apart from the determination of the statistical power of the assessment, the population under study should also be scrutinised for problems of bias and the representativeness of the disease spectrum [7]. Special attention should be given to choice or selection of the non-diseased population. For example one should ask: were these patients originally suspected of having the disease which was subsequently ruled out, or were they healthy bona fide disease-free individuals? If the distribution of test outcomes from originally suspected but eventually non-diseased patients is different from the distribution for healthy normals, this will lead to different denominator probabilities and, concomitantly, to different likelihood ratios [8]. Watson and Tang discussed the diagnostic value of a radioimmunoassay of prostatic acid phosphatase (RIA-PrAP) for the diagnosis of prostatic cancer [9]. Of the 113 patients with prostatic cancer, 79 showed a positive test outcome, 204 of 217 subjects without prostatic cancer had a negative test. Among the non-diseased subjects there were 50 normal controls and 167 with some disease, but not having prostatic cancer. All normal controls had a negative test outcome. The specificity in the total reported non-disease group is 204/217 = 0.94, therefore 13/217 were false-positive. If only normal controls were used the specificity would be 1.00. Using the total non-diseased group, the LR of a positive test would have a value of 11.7 ((79/110)/(13/217)); using only normal controls the LR-value becomes infinite!

The likelihood ratio for a simple binary test

A binary test by definition has two mutually exclusive outcomes, e.g. the presence or absence of a physical sign or of a laboratory finding. During clinical follow-up of cancer patients, the development of effusions often presents a diagnostic problem. This is a typical dyadic situation: the effusion is either present or absent. To determine the diagnostic value of a physical sign for the detection of ascites (expressed as 'present' or 'absent'), a systematic evaluation of patients suspected of having ascites with careful clinical follow-up is needed [10]. This should lead to the determination for each individual patient of whether or not he/she actually has ascites and whether a fluid-wave sign is present or absent. In Table 2, the results of such a study are summarised.

The sensitivity of the fluid-wave sign is low (0.62) and the specificity high (0.90). For each of the two test outcomes a likelihood ratio can be derived: (1) the LR for a positive sign LR(+), which is equal to 0.62/0.10 = 6.2, (2) the LR for a negative or absent sign LR(−), being 0.38/0.90 = 0.42. The likelihood ratio attaches to each of the two possible outcomes one measure expressing the discriminating power of this outcome.

The LR(−) of 0.42 indicates that it is hazardous to rule ascites out with this clinical test alone. Several clinical signs have been described for the detection of ascites. Williams and Simel used the LR-method to compare which clinical signs were most appropriate to determine the presence or absence of ascites [10].

In general, for tests with a binary outcome, such as in this example, it is easily seen from the 4-fold table and the general LR-formula that:

$$LR(+) = \frac{\text{sensitivity}}{1 - \text{specificity}}$$

and

$$LR(-) = \frac{1 - \text{sensitivity}}{\text{specificity}}.$$

Table 2. Assessment of the diagnostic value of the fluid-wave sign in patients suspected of having ascites (in parentheses, the probability of that outcome, given the final disease state is present). Adapted from Williams (see [10]*)*

	Sign (+)	Sign (−)	Total:
Final outcome			
Ascites	116 (0.62)	71 (0.38)	187 (1.00)
No ascites	13 (0.10)	117 (0.90)	130 (1.00)
LR	6.2	0.42	
with 95% CI	(3.6–10.5)	(0.34–0.51)	

Sensitivity = proportion of diseased people who have a ⊕ test = 116/187 = 0.62. Specificity = proportion of non-diseased people who have a ⊖ test = 117/130 = 0.90. LR(+) = sens/(1-spec) = 0.62/(1-0.90) = 6.2 LR(−) = (1-sens)/spec = (1-0.38)/0.90 = 0.42 .

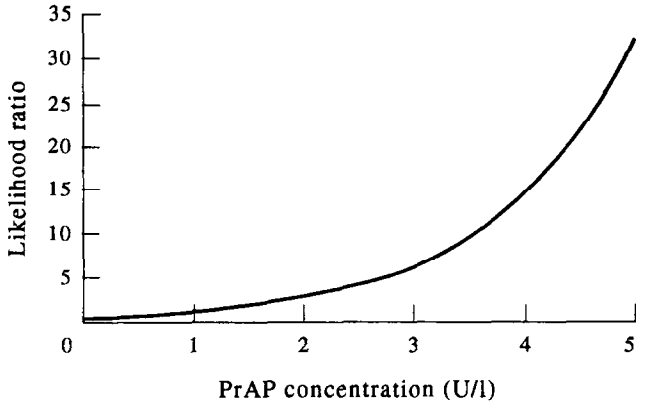

Figure 1. Likelihood ratio as a function of the PrAP-test result. This function was derived with logistic regression analysis from results adapted from Zwetsloot-Schonk and coworkers [12]:

$$LR_{CONC} = e^{(-0.699+0.833(CONC))}$$

where CONC is the concentration of PrAP in the serum, expressed in units per litre.

The likelihood ratios of tests are increasingly presented in textbooks. Sox and associates present a comprehensive list of tests with these characteristics and the publications from which they were derived [11].

The likelihood ratio for the continuous scale: clinical chemistry

By their very nature, results from the clinical chemistry laboratory are invariably quantitative and expressed on a continuous scale. To dichotomise outcomes into the categories 'normal' and 'abnormal' implies substantial loss of information: the clinical interpretation often depends on the degree of abnormality (see the discussion above about ESR in diagnosing Hodgkin's disease). For the detection of prostatic carcinoma, the prostatic acid phosphatase (PrAP) is frequently used. Zwetsloot-Schonk and coworkers studied its diagnostic value in a group of patients with a malignant prostatic disease and a group with a benign condition. To be properly informed about the diagnostic value, the frequency distribution of PrAP in both groups of patients must be known. From this 'past-experience' knowledge, the LR for different test outcomes can be estimated and used for future patients [12].

To determine the LR it is sometimes much practical to use discrete ranges of PrAP concentrations similar to Table 1. The LR for prostatic cancer of a particular range of PrAP can then be derived using the formula:

$$LR = \frac{P(\text{range}|\text{prostatic})}{P(\text{range}|\text{benign disease})}$$

Another possibility is to model the likelihood ratio as an exponential function of the test result: the higher the level of PrAP, the more likely the presence of prostatic cancer. To find the coefficients of that function, logistic regression analysis was applied to these results. The relation beween the LR and PrAP-concentration established by this technique is:

$$LR(PrAP) = EXP[-0.699 + 0.833 \times PrAP].$$

The graphic representation of this formula is given as a nomogram in Figure 1.

These kinds of studies provide either a chart similar to Table 1, or an equation as given above with a nomogram like Figure 1, which the clinician can use to appraise the test outcome.

In histo- and cytopathology, quantitative methods are increasingly applied. The numerical expression of morphological information permits the application of statistical methods for the analysis of results and likewise the LR-concept can be employed. Studying their diagnostic value for breast carcinoma, Beerman and coworkers recently reported to such variables: the mean and standard deviation of 50 cell areas measured in one section from tumour tissue [13]. These two variables can be judged simultaneously with respect to their diagnostic information and visualised in a scatterplot for 'proven' malignant and benign cases. When investigating a new patient, the two variables can be measured, and comparison of their outcomes with the distribution of the proven cases again gives rise to a LR estimate for malignancy. The technicalities for this situation are outside the scope of this paper (they can be found in [5]), but the basic principle of test appraisal is the same.

The likelihood ratio for the ordinal scale: imaging techniques and cytology

Ultimately, results from imaging techniques, cytology and histopathology are not a machine-generated quantitative result, but a conclusion based on human judgment. Customarily, results are expressed on an ordinal scale with a hierarchical order among outcome categories. When reporting, for example, a mammography, the radiologist may use one of the following classifications:

Class 1-Benign lesion (with certainty)

Class 2-Probably benign, but not absolutely certain

Class 3-Suspicious for malignancy

Class 4-Definitively malignant.

Systematic evaluation of 361 consecutive mammographies, relating test outcome to the final diagnosis established either by histological examination of the lesion or with careful clinical follow-up, may yield the results shown in Table 3.

In this contigency table, the probability of each diagnostic class, given the presence or absence of breast cancer can be calculated and used for the computation of the LR, which is displayed in the third row. Based on the individual LR, each outcome class is now fully interpretable with respect to its discriminating power between malignant and benign histology. When measures such as sensitivity and specificity are applied, appreciation of each class is virtually impossible. The usefulness of the LR-concept for the evaluation of cytology was demonstrated in a recent methodological study of fine-needle aspiration cytology (FNAC) of the breast [13]. The results of FNAC usually are expressed in four classes ranging from definitely benign to definitely malignant, including an outcome category 'unsatisfactory', which means that the smear contains too few epithelial cells, which makes it uninterpretable. In Table 4 the results of the study are presented. When examining the LR of each class, the outcome category 'definitely malignant' for example,

Table 3. Relating mammographic classification (class 1 to class 4) to final histological classification (cancer or benign). In parentheses is the probability of that outcome class given the final outcome. The estimation of the likelihood ratio (LR) for every class is accompanied by the 95% confidence intervals (95% CI)

	Class 1	Class 2	Class 3	Class 4	Total
Cancer	12 (0.09)	13 (0.10)	22 (0.16)	87 (0.65)	134 (1.00)
Benign	149 (0.66)	60 (0.26)	16 (0.07)	2 (0.01)	227 (1.00)
LR with 95% CI	0.14	0.38	2.29	65.00	
	0.08–0.24	0.21–0.64	1.27–4.28	184.–295	

shows an infinite LR and the diagnosis of cancer is thereby firmly established. Contrary to common belief, the category 'unsatisfactory', appears to contain relevant diagnostic information (LR $\neq$ 1)!

The LR-approach, as used to evaluate the accuracy of diagnostic cytology, another area of application: quality control. Since every outcome category is appreciated individually, this makes it possible to compare results from different laboratories or different observers. A study comparing the results published in the medical literature on fine-needle aspiration cytology of the breast, showed striking differences of the LR for defined outcome categories between laboratories [14]. The LR for the outcome 'malignant' ranged from 777–∞, for 'suspect' from 51–∞, for 'benign' from 0–0.31 and for 'unsatisfactory' from 0–1.09. The LR-concept is also very useful for the evaluation of cervical cytology [15]. In all these examples, the same basic principle was applied, revealing the diagnostic value of each separate outcome class.

For the full interpretation of a test result, this characterisation with the LR is needed together with the probability of disease before the test is applied. The mathematical relation between these is given with Bayes' theorem, which will be demonstrated in the next section.

LIKELIHOOD RATIO AND POST-TEST PROBABILITIES

An effective test will show a difference between the probability of disease before testing (also called prior or pretest probability) and after (post-test or revised probability). With the new information from testing, there is hopefully a difference between prior and post-test probability. This is the gain in information from testing.

The probability of the presence or absence of disease given a particular test outcome **u**, the post-test probability, can be calculated applying Bayes' theorem (for a more detailed discussion of Bayes' theorem, the reader is referred to [11], Chapter 4). The post-test probability of disease depends on the prior probability of disease **p** and the likelihood ratio of the outcome u (LR(u)) for that particular diagnostic class **D**. It is given by the following equations for the probability of the presence (P[D | u]) or absence (P[non-D |]) of disease given test outcome **u**, respectively:

$$\mathrm{P[D|u]} = \frac{p \times \mathrm{LR(u)}}{(1-p) + p \times \mathrm{LR}(u)}$$

$$\mathrm{P[non-D|u]} = \frac{1-p}{(1-p) + p \times \mathrm{LR}(u)}.$$

In Figure 2 the relation between post-test and the prior probability is depicted for several different values of the LR.

This figures shows that for LRs > 1, the higher the LR, the higher the gain in information. Similarly, for LRs < 1, the closer the LR is to zero the higher the gain in information.

These post-test probabilities given a test outcome u are also called predictive vaues. Especially for a binary outcome (positive or negative), the terminology of positive predictive value and negative predictive value is in use for the two post-test probabilities. Using the earlier mentioned expression for the likelihood ratio of a positive test outcome

$$\mathrm{LR}(+) = \frac{\text{sensitivity}}{1 - \text{specificity}};$$

the above-mentioned equation P(D | u) turns out to be identical to

$$P(\mathrm{D}|+) = \frac{\mathrm{p} * \text{sensitivity}}{(1-\mathrm{p})(1-\text{specificity}) + \mathrm{p} * \text{sensitivity}}.$$

This equation for the positive predictive value shows clearly the dependency of this value on the pretest or prior probability.

Within a group of patients tested, a variety of prior probabilities occur. This spectrum may differ from clinical situation to situation. Extreme low priors are encountered in screening asymptomatic patients for cancer. Low priors on malignancies occur often in general practice. For male

Table 4. Relation between the result of FNAC of the breast and final diagnosis. In parentheses are shown proportion of aspirations belonging to that particular FNAC category from the total given that final diagnosis. Likelihood ratios (LR) and their 95% confidence intervals (CI) are calculated

	Definitely malignant	Suspicious	Benign	Unsatisfactory	Total
Cancer	216	66	30	15	327
	(0.661)	(0.202)	(0.092)	(0.046)	(1.000)
Benign	0	27	541	71	639
	(0.000)	(0.042)	(0.847)	(0.111)	(1.000)
LR with 95% CI	∞	4.81	0.11	0.41	
	-	3.12–7.32	0.08–0.15	0.24–0.71	

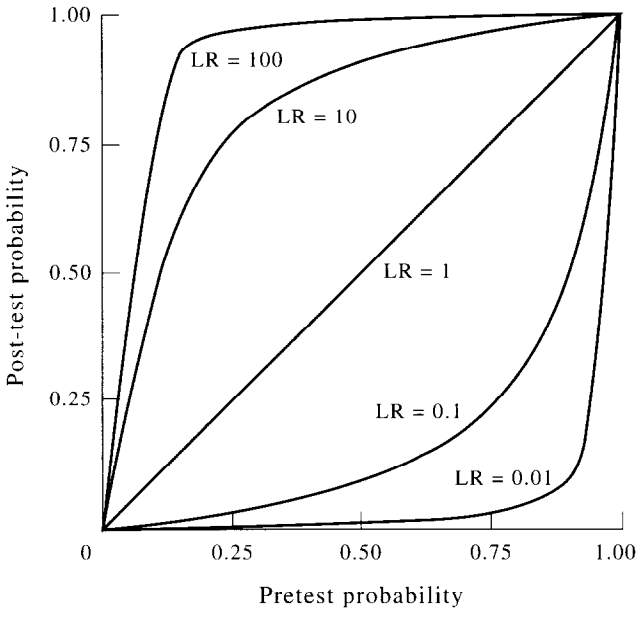

Figure 2. The relation between prior and post-test probability of disease for tests with different LR values (100, 10, 0.1 and 0.01, respectively).

patients, over 65 years of age, consulting their general practitioner with urinary tract complaints, the likelihood of having prostatic carcinoma as the cause of these complaints is much lower than in symptomatic patients referred to an outpatient urology clinic. If a firm prostatic nodule is detected during subsequent rectal examination, the probability of prostatic carcinoma may exceed 50%. A particular test outcome (e.g. 4.80 U/l) of PrAP with a LR = 11.58 for prostatic cancer (see Figure 1) leads to a very different post-test probabilities for malignancy. For the priors 0.0001, 0.01 or 0.50 we find for the post-test probabilities 0.0012, 0.105 and 0.92, respectively. So finally the diagnostic value of a test is the result of the LR and the pretest probability.

Another oncological example is the interpretation of a test result during follow-up. This interpretation depends on the prognostic category of the patient. For example, a positive bone scan in a breast cancer patient has a different meaning in a patient with no positive lymph nodes than with five axillary lymph node metastases.

Diagnostic tests are often used in sequences. A positive result from one test may precipitate a second test. Provided the two tests are independent, then in the calculation of the post-test probability from the result of the second test, the post-test probability of the first test is used as the pretest probability! As mentioned above, the LR facilitates the interpretation of simultaneously performed tests.

CONCLUSION

To achieve the objective of optimal diagnosis in patients suspected of having cancer, the primary issue is to know the possibilities and limitations of the diagnostic tests and assessment of the clinical situation where they are employed. The LR of an outcome class of a particular test gives complete information to be used for test selection and interpretation. Once this technique is mastered, it is much more appropriate than the use of sensitivity and specificity.

Recently, a special working group advocated the use of numerical data, derived from systematic analysis of daily practice, to exercise evidence-based medicine [16]. A good book on the statistical issues connected to evidence-based medicine appeared recently [17]. It is remarkable that it is hard to find methodologically sound examples of test-evaluation in the medical literature. When reviewing all articles published on the evaluation of FNAC of the breast, most publications lacked the essential data on selection of patients and demographic data, the disease spectrum etc. as described by Sackett and colleagues to assess their validity [3, 18]. It is far from common practice to handle diagnostic information at a high quantitative level. Often a LR or a pretest probability are not accurately known. However, their magnitude can be estimated.

It should be realised that this technique for evaluation provides no less than a starting point in the process of rational medical decision-making. Another important issue is to ascertain how certain we need to be about a diagnosis. This is the domain of medical decision analysis.

1. Kasssirer JP. Our stubborn quest for diagnostic certainty. A cause of excessive testing. *N Engl J Med* 1989, **320**, 1489–1491.
2. Sox HC. Probability theory in the use of diagnostic tests. An introduction to critical study of the literature. *Ann Int Med* 1986, **104**, 60–66.
3. Giard RWM, Hermans J. The interpretation of diagnostic cytology with likelihood ratios. *Arch Pathol Lab Med* 1990, **114**, 852–854.
4. Fletcher RH, Fletcher SW, Wagner EH. Clinical epidemiology: the essentials. Chapter 2: Abnormality. Baltimore, Williams & Wilkins, 1988, 19–42.
5. Albert A, Harris EK. Multivariate interpretation of clinical laboratory data. New York, Marcel Dekker Inc., 1987, 90–91.
6. Centor RM. Estimating confidence intervals of likelihood ratios. *Med Decis Making* 1992, **12**, 229–231.
7. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1987, **299**, 926–930.
8. Lachs MS, Nachamkin I, Edelstein PH, *et al.* Spectrum bias in the evaluation of diagnostic tests: lessons from the rapid Dipstick Test for urinary tract infection. *Am Int Med* 1992, **117**, 135–140.
9. Watson RA, Tang DB. The predictive value of prostatic acid phosphatase as a screening test for prostatic cancer. *N Engl J Med* 1980, **303**, 497–499.
10. Williams JW, Simel DL. Does this patient have ascites? How to divine fluid in the abdomen. *JAMA* 1992, **267**, 2645–2648.
11. Sox HC, Blatt MA, Higgins MC, Marton KI. *Medical Decision Making.* Boston, Butterworths, 1988.
12. Zwetsloot-Schonk JHM, Schmitz PIM, Hermans J, Frölich M. What clinicians should know when using Bayes theorem: prostatic acid phosphatase test for detecting prostatic cancer. *Theor Surg* 1992, **7**, 117–123.
13. Beerman H, Veldhuizen RW, Blok APR, *et al.* Cytomorphometry as quality control for fine needle aspiration: a study in 321 breast lesions. *Analyt Quant Cytol Histol* 1991, **13**, 143.
14. Giard RWM, Hermans J. The value of aspiration cytologic examination of the breast. A statistical review of the medical literature. *Cancer* 1992, **69**, 2104–2110.

15. Giard RWM, Hermans J. The evaluation and interpretation of cervical cytology with likelihood ratios. *Cytopathology* 1993, **4**, 131–137.
16. Evidence-based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA* 1993 **268,** 2420–2425.
17. Dunn G, Everitt B. *Clinical Biostatistics. An Introduction to Evidence-Based Medicine.* London, Edward Arnold, 1995.
18. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. How to read a clinical journal. In *Clinical Epidemiology. A Basic Science for Clinical Medicine*, 2nd edn, Chptr 12. Boston, Little, Brown & Co., 1991, 285–322.