European Journal of Epidemiology **10**: 29–36, 1994.

Likelihood ratios: A real improvement for clinical decision making?

Bruno Dujardin[1], Jef Van den Ende[2,4], Alfons Van Gompel[2,4], Jean-Pierre Unger[1] & Patrick Van der Stuyft[3]

[1] *Unit for Research and Training in Public Health;* [2] *Clinical Department, and* [3] *Epidemiology Unit, Prince Leopold Institute for Tropical Medicine, Antwerp, Belgium;* [4] *Unit for Tropical Diseases, University Hospital, Edegem, Antwerp, Belgium*

Accepted in revised form 22 December 1993

Abstract. The concept of likelihood ratio has been advocated for several years as one of the better means to evaluate diagnostic tests and as a practical and valuable tool in clinical decision making. In this paper we review the basic concepts underlying the evaluation of diagnostic tests and we explore the properties and usefulness of both positive and negative likelihood ratios compared with sensitivity and specificity. Particular attention is given to the use of likelihood ratios in the clinical setting. Likelihood ratios have three main advantages: they are intuitive, they simplify the predictive value calculation and the overall evaluation of sequential testing. Disadvantages are the non-linearity and the necessity to recalculate probabilities in odds. Although they summarize the information contained in sensitivity and specificity, these characteristics are still necessary for certain clinical decisions. Since likelihood ratios have been promoted among physicians and medical students, we discuss examples of inappropriate use and misunderstandings in the medical literature: the frequent omission of confidence intervals, the choice of cut-off points based on likelihood ratios for positive test results only and the confusion between likelihood ratios for ranges and those for cut-off points.

Key words: Diagnostic test evaluation, Likelihood ratios

Introduction

Until recently, the few publications addressing the issue of likelihood ratios were restricted to radiology and clinical chemistry. However, over the last 10 years several clinical epidemiologists have proposed their use for evaluation of diagnostic tests in hospital units [1–12]. Some authors are very enthusiastic: 'It is recommended that clinical laboratories improve displays of data by replacing reports of sensitivities and specificities with likelihood ratios calculated for multiple levels of tests results' [11], or: 'The likelihood ratio plays a central role in probability calculations and in the decision analysis' [8]. In 1989, the *Diagnostic testing handbook for clinical decision making* was published in which more than 140 diagnostic tests with their sensitivity, specificity and likelihood ratios are described [5]. As stated by the editors: 'This manual was designed to provide an easy to use listing of diagnostic tests for medical students, house officers and practicing physicians'.

Is this preference for likelihood ratios justified? In general, what are the criteria that make a test useful for decision making in clinical practice? In the case of a quantitative test, what are the differences between likelihood ratios for ranges or for cut-off points? What are the advantages and disadvantages of positive (LHR+) and negative (LHR–) likelihood ratios compared with sensitivity and specificity? This paper is intended to review the properties of likelihood ratios as a tool for clinical decision making, their advantages and drawbacks and some common misuses and misunderstandings. For a complete evaluation of decision analysis, other tools are needed: decision trees and utility estimation [13].

Basic definitions

Example

We first present a brief overview of classical concepts, applied to an example, the LD1/LD2 ratio test (Lactate Dehydrogenase, EC 1.1.1.27), one of the cardiac enzymes used in hospitals to diagnose myocardial infarction in patients presenting with chest pain [10]. We refer to standard texts for an in depth treatment [8, 12]. Figure 1 shows the distribution of LD1/LD2 values for 97 cases of myocardial infarction and 146 patients with chest pain but without infarction. The diagnosis was established by a cardiologist on the basis of both electrocardiographic and scintigraphic results (gold standard).

Table 1 shows the distribution of cases and non-

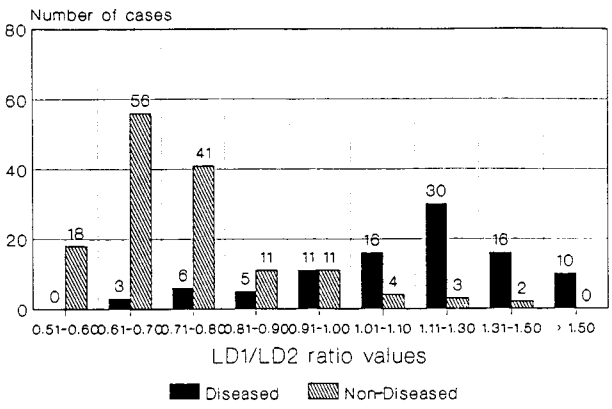

Figure 1. Distribution of LD1/LD2 ratio values in diseased and non-diseased patients.

Table 1. Contingency table for the decision level: LD1/LD2 >1.00

		DISEASE		
		Present	Absent	
TEST RESULT	> 1.00	72 (TP)	9 (FP)	81
	< 1.00	25 (FN)	137 (TN)	162
		97	146	243

cases when a decision level (or 'cut-off point') LD1/LD2 greater than 1.00 is chosen for the LD1/LD2 ratio. With this procedure, continuous data have been forced into a binary classification. As we shall see later, some diagnostic information is lost.

Sensitivity, specificity and predictive values

The result of any test can be interpreted as an argument to strengthen or to weaken a disease hypothesis based on the available information on the patient. The strength of the argument will be translated in positive and negative predictive values. The positive predictive value (PPV) is defined as the probability of disease when the test result is positive.

PPV = TP/(TP+FP)
(72/81 = 89% for the LD1/LD2 ratio test).

The negative predictive value (NPV) is the probability of absence of that particular disease if the test result is negative.

NPV = TN/(TN+FN)
(137/162 = 85% for the LD1/LD2 ratio test).

PPV and NPV will be determined by the combination of the sensitivity and the specificity values of the test for a given disease, and by disease prevalence (Pr). Sensitivity (Se) is the proportion of diseased with a positive test result: Se = TP/(TP+FN). In our example sensitivity was 74% (72/97). Specificity (Sp) is the proportion of non-diseased with a negative test result (Sp = TN/(TN + FP)), 94% in our example.

By definition, sensitivity and specificity can be evaluated only if the real health status is known: diseased or non-diseased. While a binary test has only one value for Se and Sp, Se and Sp will vary according to the chosen decision level for a quantitative test (ordinal or continuous). Patients will shift from test positive to test negative with a changing cut-off point (Figure 1) and for this reason, Se and Sp are not independent but inversely related (4).

For the individual patient, the prevalence rate of the disease is represented by the pretest probability of having the disease. In most cases, the real prevalence rate in the considered population is not known and clinicians have to rely upon an estimate based solely on their own experience. Bayes' theorem gives the relationships between PPV, NPV and Se, Sp, Pr [14]:

$$PPV = \frac{Pr * Se}{Pr * Se + (1 - Pr) * (1 - Sp)}$$

or

PPV = TP/(TP + FP)

$$NPV = \frac{(1 - Pr) * Sp}{(1 - Pr) * Sp + Pr * (1 - Se)}$$

or

NPV = TN/(TN + FN)

Thus, when the prevalence rate (or pretest probability) increases, PPV increases. In the past, predictive values for an estimated prevalence rate of 50% , or worse, predictive values based on the study populations, were often given as a standard characteristic for a test. These 'standard' predictive values are of course useless in other settings where the prevalence rate or pretest probability is different. Considerable confusion has been the consequence, and clinicians continue to misunderstand this issue.

Likelihood ratios

Calculation of PPV and NPV with Se and Sp is quite time consuming. Moreover, the strength of a positive or a negative test result has to be given in two numbers which have to be combined: this is not so evident for the clinician. An alternative approach is given by the likelihood ratio (LHR).

Likelihood ratios are ratios of two probabilities (or likelihoods) [3, 11, 12]:

$$LHR = \frac{\text{Probability of test outcome given diseased patients}}{\text{Probability of test outcome given non-diseased patients}}$$

They may be calculated in two different ways,

depending on the nature of the test used: (1) In the case of a qualitative (dichotomous) test or when the data of a quantitative test are expressed dichotomously for each decision level (or cut-off point), there are only two likelihood ratios, one for a positive test result (LHR+) and one for a negative test result (LHR–):

LHR+ = Se/(1-Specificity) *or*
LHR+ = TP rate/FP rate

and

LHR– = (1-Sensitivity)/Sp *or*
LHR– = FN rate/TN rate

In our example in Table 1 (see also Table 2b), the LHR+ = (72/97)/(9/146) = 12.0. This means that an LD1/LD2 ratio above 1.00 is twelve times more likely to occur in a patient with myocardial infarction than in a patient with chest pain but without infarction. The LHR– has a value of (25/97)/(137/146) = 0.27, showing that an LD1/LD2 ratio equal to or below 1 is less than three tenths as likely to be found in a patient with a myocardial infarction as in a patient without infarction.

(2) In case of quantitative tests, likelihood ratios may be calculated for each outcome level or range (i) of a diagnostic test result. The likelihood ratio for a given range (LHRi) is the ratio of the probability of a positive (abnormal) test result when the disease is present divided by the probability of the same test result when the disease is absent:

$$LHR_i = \frac{TP_i rate}{FP_i rate}$$

The likelihood ratio at range (i) expresses how many times more likely an abnormal test result with a value included in this range (i), is to be expected in diseased patients as compared with non-diseased patients (i.e. patients not suffering from the specific disease under investigation). It is important to distinguish between LHRi for ranges or slices ('i'), where only the numbers between two values of test results are taken into account, and likelihood ratios for dichotomous tests (see next paragraph), where cumulative numbers above and below the cut-off point are used.

Table 2a shows, for each of the nine different LD1/LD2 value ranges, the total of true positive and false positive test results, and the corresponding values of true positive rates, false positive rates, and the likelihood ratio for the different ranges. Table 2b shows for nine cut-off points (the upper value of each

Table 2a. Absolute numbers of True Positives (TP), False Positives (FP), True Positive rates, False Positive rates and LHRxi (with 95% confidence intervals) for each xi range of LD1/LD2 ratio values

No.	Ranges	TP (n)	FP (n)	TP (n)	FP (n)	LHRi	CI (95%)
9	> 1.50	10	0	10	0	–	
8	1.31–1.50	16	2	16	1	12.0	2.8–51.2
7	1.11–1.30	30	3	31	2	15.1	4.7–48.0
6	1.01–1.10	16	4	16	3	6.0	2.1–17.5
5	0.91–1.00	11	11	11	8	1.5	0.7–3.3
4	0.81–0.90	5	11	5	8	0.7	0.2–1.9
3	0.71–0.80	6	41	6	28	0.2	0.1–0.5
2	0.61–0.70	3	56	3	38	0.1	0.0–0.3
1	0.51–0.60	0	18	0	12	0	0
	Total	97	146				

Table 2b. Absolute numbers of True Positives (TP), False Negatives (FN), Sensitivity (%), Specificity (%), LHR+ and LHR– (with their 95% confidence interval) for each cut-off point (= upper value of each xi range of LD1/LD2 ratio values)

Decision level	TP (n)	TN (n)	Sensitivity (CI-95%)	Specificity (%)	LHR+ (CI-95%)	LHR–
> 1.50	10	146	10 (4–16)	100	–	0.90
> 1.30	26	144	27 (18–36)	99	19.6 (4.8–80.6)	0.74
> 1.10	56	141	58 (48–68)	97	16.9 (7.0–4.06)	0.44
> 1.00	72	137	74 (65–83)	94	12.0 (6.3–22.9)	0.27
> 0.90	83	126	86 (79–93)	86	6.2 (4.1–9.5)	0.17
>0.80	88	115	91 (85–97)	79	4.3 (3.1–5.9)	0.12
> 0.70	94	74	97 (93–100)	51	2.0 (1.7–2.3)	0.06
> 0.60	97	18	100	12	1.1 (1.0–1.2)	0.00
> 0.50	97	0	100	0	1	

range has been chosen as the cut-off point), true positive (TP) and true negative (TN) cases, sensitivity and specificity and the LHR+ and LHR–.

For example, at the seventh highest range, LHR_7 = (30/97)/(3/146) = 15.1 (Table 2a).

On the other hand, taking 1.3 as the cut-off point (Table 2b), counting all values equal or superior to 1.3 as positive, Se = 26/97 = 27% and Sp = 144/146 = 99%. LHR+ would be 19.6 (26/97)/(2/146) and LHR– = 0.74 (71/97)/(144/146). As Se and Sp, LHR+ and LHR– are inversely related. The absolute value of LHR– however, will increase with diminishing discriminative power of the test.

LHRi and the LHR+ may vary from 0 to infinity [7]. The higher the value, the more predictive a positive test will be for a given prevalence. An LHRi (or an LHR+) of 1 means that the test has no value: an abnormal result is as likely to be found in diseased as in non-diseased patients. Note that LHRi and LHR+ may be interpreted as a cost-benefit ratio: the numerator or true positive rate represents a benefit criterion, whereas the denominator or false positive rate stands for cost, in that false positives will cause unnecessary further testing, inappropriate treatment or anxiety.

The negative likelihood ratio (LHR–) is the ratio of the probability of a normal test result in diseased patients (false negative fraction) divided by the probability of a normal test result when the disease is absent (true negative fraction). The LHR– expresses how many times less likely a normal test result is to be expected in diseased patients as compared to non-diseased. Its clinically interesting values vary between 0 and 1, and the smaller the value, the higher the negative predictive value for a given prevalence. Here also, the LHR– can be seen as a cost/benefit ratio. The false negative rate represents cost, the true negative rate the benefit. Note that the LHR– does not exist for tests with results in more than 2 ranges (or slices): with several ranges, a negative test result represents another range itself. In this case, a test result outside a given range means that it must lie in one of the other ranges, and cannot be named 'negative'. It makes no sense to calculate the value of a test result 'provided it is outside this range, either higher or lower' [000].

Test characteristics

Sensitivity, specificity: advantages

1. Se and Sp of a test do not depend on disease prevalence.
2. For the decision making process, a diagnostic test with a higher Se is preferred when 'false negatives' are to be avoided, for example if the suspected disease is life-threatening and an effective and cheap, non-toxic treatment exists (e.g. fever for suspected malaria). A test with high specificity has to be chosen when false positives should be avoided, for example an untreatable disease or a dangerous treatment (e.g. histology for cancer chemotherapy). A high sensitivity is necessary to definitely 'rule out' a diagnosis (i.e. make sure a patient has not the disease), a high specificity to 'rule in'.

Se, sp: disadvantages

1. Sensitivity only applies to patients with a particular disease, and specificity only to healthy people, or to patients without that particular disease. A clinician will not start from diseased or not diseased, but from a positive or a negative test. Therefore, Se and Sp are intuitively not so evident as LHR in the routine clinical environment.
2. Both values have to be combined, using Bayes theorem, to have an idea about the strength of a positive or a negative test.
3. Although Se and Sp of a test do not depend on prevalence, they do depend on variables such as disease stage [15], age and gender of the patient and characteristics of the different settings the test is used in [11]. This is the reason why it is sometimes (erroneously) stated that Se and Sp can vary with prevalence.

LHR: advantages [2, 3, 7, 8, 11, 12, 14]

1. Likelihood ratios (positive PLUS negative) *summarize the information* of both sensitivity and specificity and give the discriminative power of the test. A likelihood ratio of 1.0 implies that the test is of no value: the same proportion of diseased and non-diseased patients have the same test result. An LHR+ below 1 or an LHR– above 1 implies that Se + Sp < 1. In this situation, more patients are classified incorrectly after the test than before.
2. Like sensitivity and specificity, likelihood ratios *don't vary with changes in prevalence*. As stated before, this should not be misunderstood: Se, Sp, likelihood ratios and prevalence all depend on characteristics of different settings, but Se, Sp and LHR do not depend on prevalence itself.
3. If the diagnostic test has several levels of results, different LHRi may be calculated for each range of values, and not only for cut-off points. For Se and Sp we are limited to a single chosen threshold. LHRi thus efficiently increase the clinical information content of a quantitative diagnostic test and provide *a more independent characteristic* than Se and Sp: as the setting varies, the mix of more versus less serious cases can vary, and with this the Se and Sp for a given cut-off point vary. LHRi however are relatively

independent, as serious and less serious cases will tend to show up in their corresponding range 'i'. As seen in Table 1, we have only one decision level with an LHR+ of 12.0 for all test values equal or superior to 1.00. This LHR+ is an average of the different LHRi values: an LHRi of 6.0 (16/97)/(4/146) for an LD1/LD2 ratio value between 1.01–1.10 and LHRi of 15.1 and 12.0 for LD1/LD2 ratio values between 1.11–1.30 and between 1.31–1.50. It is obvious that the discriminative power of an LD1/LD2 test result between 1.01–1.10 is quite different from an LD1/LD2 value between 1.11–1.30, although both are considered 'positive'. Thus, when the results of a quantitative test are expressed dichotomously, the LHR+ is a mix of different LHRi values. High range LHRi values may, however, be more common in one setting than in another. If an LHR+ is calculated from LHRi values observed in a hospital setting and then applied in a general practice, the discriminant power of a positive test may be overestimated.

4. If a disease's pretest probability (or prevalence) is known or can be estimated, likelihood ratios allow for *direct calculation of posttest probability* (predictive values) using a formula which can be easily derived from Bayes theorem [16].

   Pretest odds * Likelihood ratio = Posttest odds

   with Pretest odds = Prevalence/(1 – Prevalence) and Predictive value = Posttest odds/(1 + Posttest odds). This is the major advantage of LHR, and gives it superiority over the 'predictive values for given prevalence', which have to be calculated by the rather complicated Bayes Theorem.
5. If a *serial testing* strategy is used, a final predictive value can easily be calculated multiplying the posttest odds of the previous test with the likelihood ratio of the following one [3]. The discriminative power of a cluster of tests can also be calculated by simple multiplication of individual likelihood ratios. A condition however, has to be fulfilled: the tests shoud be totally independent [3, 17], which is seldom the case in reality. Otherwise cumulative discriminative power and posttest probability will be overestimated.

LHR: disadvantages

1. Although LHR+ and LHR– together contain all the information given by Se and Sp and are sufficient for most clinical decisions, *Se and Sp are still necessary* when false positives or false negatives have to be avoided as much as possible. The same LHR+ can be the result of the combination of very different values for Se and Sp: an LHR+ of 10 can result from the combination of a sensitivity and a specificity of, respectively, 10 and 99, 40 and 96, 80 and 92. Corresponding LHR– values (1 – Se/Sp) would be 90/99 = 0.91, 60/96 = 0.60 and 20/92 = 0.21. These values differ considerably. Although Se and Sp could be recalculated from both LHR+ and LHR–, this is a time consuming and difficult job.
2. The need to *convert back and forth between pretest probability/pretest odds* and posttest odds/predictive value may be confusing, but fortunately Fagan's nomogram [16] obviates all the calculations. Using a ruler, this graph makes it possible to find predictive values, once prevalence and likelihood ratios are known. Alternatively clinicians can be taught to think in terms of odds, instead of percentages, since these terms are used in everyday conversation [3, 13, 18].
3. *LHR values are not linear*, as they are the result of division and not of subtraction. This has important implications. Intuitively, the clinician rates the discriminative strength of an LHR+ of 100 as ten times that of an LHR+ of 10, which is an overestimation. For a pretest probability of 0.01, which is equivalent to a pretest odds of 0.01/0.99 or 0.01, the posttest odds would be 0.1 or 1 with an LHR+ of 10 resp. 100, but the positive predictive value would be 0.09 or 0.5. For a pretest probability of 0.2 the posttest probability would vary from 0.55 to 0.71 with an LHR+ of 5 resp. 10. In fact, logarithmic values of LHR better represent the discriminative strength, but calculation of predictive values would be more difficult.
4. In the case of a quantitative test with only a few cases at the highest and/or lowest ranges of values, the *precision of LHRi for those highest and/or lowest categories* or ranges of values is low. Slight fluctuations in the number of cases may produce important changes in the LHRi value. In our example (Figure 1 and Table 2a), in 18 cases the LD1/LD2 ratio is between 1.31 and 1.50 (range N°8), with 16 true positives and 2 false positives. If by chance we had 5 false positives instead of 2, the new LHR_8 value would become (13/97)/(5/146) = 3.9 instead of 12.0.

For the diagnosis of hypothyroidism, for instance, a T4 value between 4.1–5.0 µg/l has an LHRi of 13.8 [19]. Such a high value seems very evocative, and could prompt the clinician to consider an appropriate treatment. In fact, the true likelihood ratios for the data presented lies somewhere between 1.9 and 71. Given such a result, the conclusion is that T4 testing could be a relevant diagnostic test for hypothyroidism, but more patients are needed in order to obtain a more precise estimate of the LHRi or to make the confidence interval less wide.

For Se and Sp the standard statistical procedures for constructing confidence intervals (CI) round a proportion can be used. For likelihood ratios several methods have been proposed. First, Mietinen's

formula for the estimation of the confidence interval for a risk ratio may be used [2]:

$$95\%CI = LHR_i^{1 \pm 1.96\sqrt{\chi^2}}$$

But Mietinen's formula is an approximation and other more 'exact' methods such as Simel's formula [20] have been advanced:

$$LHRi = \exp\left(\ln(P1/P2) \pm 1.96 \sqrt{\frac{1 - P1}{P1 * n1} + \frac{1 - P2}{P2 * n2}}\right)$$

where P1 = sensitivity, P2 = 1 – specificity, n1 = TP + FN, n2 = TN + FP.

Epidemiological software such as Epi Info 5.0 permits easy calculation of CI, whether by 'exact' or more approximate formula. We calculated the confidence intervals for the values of LHRi and LHR+ in Tables 2a and 2b with the Simel formula. The results are presented in Figures 2a and 2b. The 95% confidence intervals for the sensitivity are presented in Figure 2c.

As we see in Figure 2a, CI for LHRi tend to widen with higher test results indicating a lack of stability of the LHRi values at highest LD1/LD2 ratio values. Figure 2b shows the same lack of stability for LHR+ values. In contrast, for the sensitivity, the overall variation of CI is less than for likelihood ratios and the widest CI are around 50% (Figure 2c).

These wide confidence intervals are to interpreted cautiously when one takes into account the 'non-linearity' of LHR. A logarithmic transformation stabilises the variability of the LHR and their CI, but log LHR are not usually reported in the literature.

Misuse of LHR in medical literature

A Medline search on likelihood ratios from 1985 through November 1990, confirmed its widespread use in medical literature. Some problems were identified.

1. The first common inaccuracy in the use of the likelihood ratio concept is the *omission of the confidence intervals*. As discussed before, this is particularly important for the highest decision levels.
2. Another common mistake is perhaps of even greater importance, namely to choose a *cut-off point on the basis of LHR+ values only*, and not to take into account the information content of LHR–. It is as if LHR– were the inverse of LHR+, or easily calculable from LHR+. As we explained, this is not the fact, and for different tests LHR– can vary widely for the same LHR+. When choosing a suitable cut-off point one should minimize the consequences of both false positive and false negative test results [16, 21]
3. Last but not least, there is an important confusion in the understanding of likelihood ratios. Too often, it is not clear whether likelihood ratios *for cut-off points or for ranges* are concerned. Interpretation of both is quite different [22].

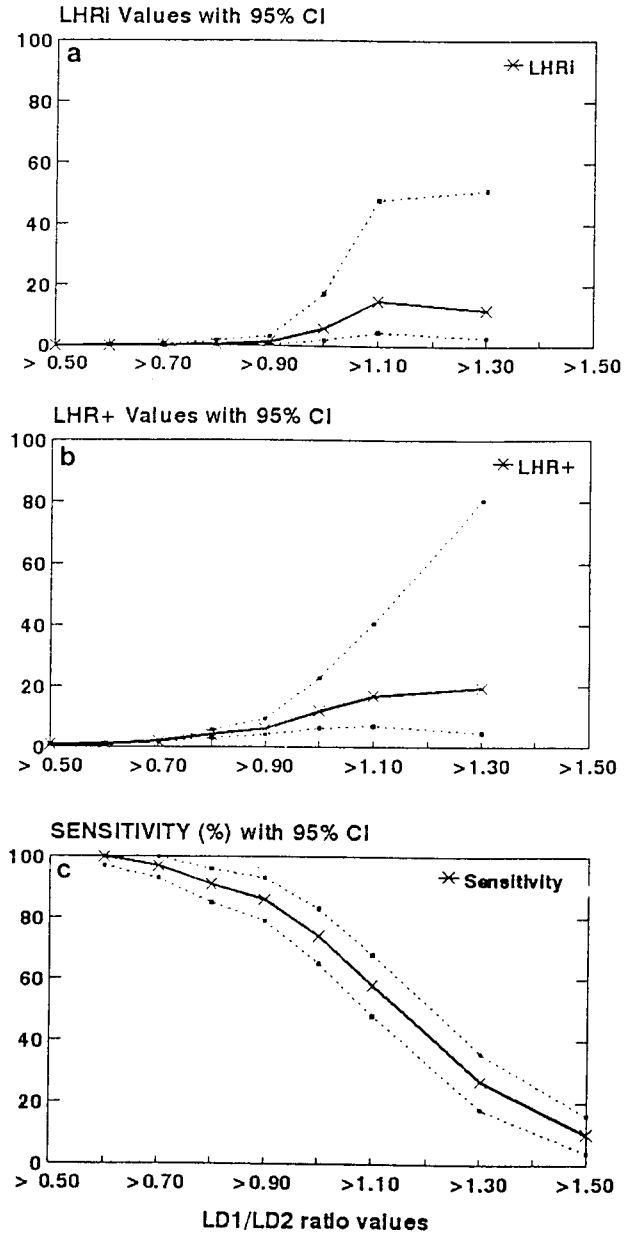

Figure 2. (a) Confidence intervals (95%) for LHRi at different ranges of LD1/LD2 ratio values; (b) Confidence intervals (95%) for LHR+ at different cut-off points of LD1/LD2 ratio values; (c) Confidence intervals (95%) for Sensitivity values at different cut-off points of LD1/LD2 ratio values.

Discussion

Five conditions for a good test characteristic can be postulated [8, 14, 21, 23–25]. It should give a correct estimate of the discriminative power of the test, be independent of prevalence, be stable in different populations, be quantifiable, and represent a tool for decision making at the bedside. We discuss here if these five conditions are fulfilled.

1. The first condition is that this characteristic should represent the *discriminative power* of a test. Neither Se nor Sp alone can provide this,

whereas, given a positive result, LHR+ and LHRi can give a reliable estimate. LHR– can give the same, given a negative result. Overall test power however, needs both LHR+ and LHR–, or all LHRi. Although discriminative power can be found from the combination of Se and Sp as well, calculation is still necessary. On the other hand, the representation of discriminative power by LHR is not linear and the precision of LHRi at high test result values tends to be lower.

2. The second condition is that the characteristic should be *independent of prevalence*. Like Se and Sp, likelihood ratios are independent of prevalence or estimated pretest probability. Up until now, predictive values are often given as test characteristics, but they are useless as they depend on disease prevalence.
3. Ideally, the characteristic's values should remain *stable in different populations*. Like Se and Sp, likelihood ratio values can change with different settings. However, calculation of likelihood ratios for ranges of test results can augment considerably the independence of test characteristics in different settings, contributing to stability.
4. The characteristic must be *quantifiable* and must allow for the calculation of predictive values. Calculation of PPV and NPV is possible if clinicians are accustomed to estimating pretest probability in terms of odds, where Fagan's nomogram could be of help. The 'predictive values for an estimated prevalence of 50%', which were used in the past, do not allow for calculation of predictive values for a given clinical situation.
5. The characteristic's *calculation should be simple*, it should be acceptable and make sense to the clinicians who will use it. When compared with classical calculation of predictive values with Bayes' formula, calculation is very simple and possible at the bedside. Even calculation of discriminative power of a cluster of tests, or predictive values for serial tests, is simple. It should be stressed however that these tests need to be independent, which is not always the case. The representation of the LHR+ as a whole number on a scale from 1 to (seldom more than) 100 makes sense even to the clinician who is unexperienced in clinical epidemiology. The intuitive understanding of LHR– however is more difficult, as is the non-linearity of both.

Conclusion

Likelihood ratios give an estimate of the discriminative power of a test or a cluster of tests, are independent of disease prevalence and represent a valuable tool for decision making at the bedside. The non-linearity, the difference between LHR for cut-off points and for ranges, and the broad confidence intervals at high test result values should be emphasized. For a dichotomous test, both positive and negative likelihood ratios have to be provided, since LHR– is not the mere mathematical inverse of LHR+. For certain clinical decisions, where it is extremely important to avoid false positives or false negatives, Se or Sp are still superior.

Acknowledgement

This research has been supported by a grant from the Fonds National de la Recherche Scientifique et Médicale, Belgium.

References

1. Albert A. On the use and computation of likelihood ratios in clinical chemistry. Clin Chem 1982; 28: 1113–1119.
2. Beck JR. Likelihood ratios: Another enhancement of sensitivity and specificity (editorial). Arch Pathol Lab Med 1986; 110: 685–686.
3. Fletcher RH, Fletcher SW, Wagner EH. Clinical epidemiology. Baltimore: Williams & Wilkins, 1988.
4. Giard RW, Hermans J. Interpretation of diagnostic cytology with likelihood ratios. Arch Pathol Lab Med 1990, 114: 852–854.
5. Goldenberg K, Verdain Barnes H, Redding MM. Diagnostic testing handbook for clinical decision making. Chicago: Yearbook Medical Publishers, 1989.
6. Guyatt GH, Patterson C, Ali M, et al. Diagnosis of iron-deficiency anemia in the elderly. Am J Med 1990; 88: 205–209.
7. Lacher DA. Special report: Predictive value derived from likelihood ratios: A superior technique to interpret quantitative laboratory results. Am J Clin Pathol 1987; 87: 673–676.
8. Linnet K. A review on the methodology for assessing diagnostic tests. Clin Chem 1988; 34: 1379–1386.
9. Pauker SG, Kopelman RI. Clinical Problem Solving. How sure is sure enough? N Engl J Med 1992; 326: 688–691.
10. Pellar TG, Leung YG, Henderson AR. A computer program for rapid generation of receiver operating characteristic curves and likelihood ratios in the evaluation of diagnostic tests. Ann Clin Biochem 1988; 25: 411–416.
11. Radack KL, Rouan G, Hedges J. The likelihood ratio An improved measure for reporting and evaluating diagnostic test results. Arch Pathol Lab Med 1986; 110: 689–693.
12. Sackett DL, Haynes RB, Tugwell P. Clinical epidemiology: A basic science for clinical medicine. Boston, Toronto: Little, Brown and Company, 1985.
13. Thornton RJ, Lilford RJ, Johnson N. Decision analysis in medicine. Br Med J 1992; 304: 1099–1103.
14. Feinstein AR. Clinical epidemiology: The architecture of clinical research. Philadelphia: Saunders, 1985.
15. Fletcher RH. Carcinoembryonic antigen. Ann of Int Med 1986; 104: 66–73.

16. Fagan TJ. Nomogram for Bayes's theorem (letter). N Engl J Med 1975; 293: 257.
17. Seymour DG, Green M, Vaz FG. Making better decisions: Construction of clinical scoring systems by the Spiegelhalter-Knill-Jones approach. Br Med J 1990; 300: 223–226.
18. Simel DL. Playing the odds. The Lancet 1985(i): 329–330.
19. Goldstein BJ, Mushlin AI. Use of a single thyroxine test to evaluate ambulatory medical patients for suspected hypothyroidism. J Gen Intern Med, 1987; 2: 20–24.
20. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: Sample size estimation for diagnostic test studies. J. Clin. Epidemiol 1991; 44 (8): 763–770.
21. Bulpitt CJ. Confidence intervals. Lancet 1987(i): 494–497.
22. Leung FY, Galbraith LV, Jablonsky G, Henderson AR. Re-evaluation of the diagnostic utility of serum total creatine kinase and creatine kinase-2 in myocardial infarction. Clin Chem 1989; 35: 1435–1440.
23. Cochrane AL, Holland WW. Validation of screening procedures. Br Med Bull 1971; 27: 3–8.
24. Linnet K. Precision of sensitivity estmations in diagnostic test evaluations: Power funçtions for comparisons of sensitivities of two tests. Clin Chem 1985; 31: 574–580.
25. Sheps SB, Schechter MT. The assessment of diagnostic tests: A survey of current medical research. JAMA 1984; 252: 2418–2422.

Address for correspondence: Dr Bruno Dujardin, Public Health Unit, Institute of Tropical Medicine, Nationalestraat 155, B-2000 Antwerpen, Belgium
Phone: (32-3) 247 6286; Fax: (32-3) 216 1431