Research Series

How to Use and Interpret Interval Likelihood Ratios

Jeffrey Sonis, MD, MPH

Background: *Likelihood ratios offer important advantages over sensitivity and specificity for characterizing diagnostic tests. They can capture the magnitude of abnormality of test results, whereas sensitivity and specificity require that the test results be dichotomized into positive or negative. This is an important advantage because many diagnostic tests are measured on continuous or ordinal scales. Posttest probabilities calculated from interval likelihood ratios may be different than those calculated from sensitivity and specificity; clinical decisions derived from the use of likelihood ratios may therefore be different from decisions derived from test results characterized by sensitivity and specificity. This article demonstrates the advantages, use, and interpretation of interval likelihood ratios using the clinical scenario of a young child with a high fever.*

(Fam Med 1999;31(5):432-7.)

Likelihood ratios are a method of characterizing diagnostic tests. They offer several important advantages compared to traditional measures such as sensitivity and specificity.[1-3] This article describes the use, interpretation, and advantages of interval likelihood ratios, using a realistic clinical example of a young child with a high fever.

Clinical Scenario

A 1-year-old boy is brought to your office with a history of high fever for 2 days. He has no other symptoms. On physical exam, his temperature is 39.5˚C, by ear. He does not appear toxic. The physical examination is normal. You are concerned about the possibility of occult bacteremia.

Diagnostic Tests and Probability of Disease

An estimate of the probability of disease, based on information available prior to diagnostic testing, is called the pretest probability. In the absence of any information about the patient other than demographic details, the pretest probability is the prevalence of disease in that population. The pretest probability of occult bacteremia in the 1-year-old boy in the clinical scenario is 4%, because the prevalence in children 3–36 months with temperature greater than 39.0˚C is 4%.[4]

A diagnostic test can be defined as a procedure that changes the estimate of the probability of disease. The diagnostic test provides information that permits the estimate of the probability to be revised. This revised estimate is called the posttest probability.

The purpose of performing a diagnostic test is to increase the estimate of probability high enough to begin treating the patient, or decrease the estimate low enough to consider the disease ruled out.[5] Some have recommended using a white blood cell (WBC) count in the evaluation of a young child with fever without obvious source.[4] The purpose of performing a WBC count is to determine whether the probability of occult bacteremia is high enough to begin treatment with antibiotics or low enough to stop testing.

Sensitivity, Specificity, and Probability of Disease

Sensitivity is defined as the probability of a positive test in those with the disease.[1] In tests with only two outcomes (positive/negative), calculation of sensitivity is straightforward. When test results are measured on a continuous scale, or there are multiple ordered outcomes, a cutoff point must be selected to characterize results as positive or negative. Most studies of fever in children have chosen a cutoff point of 15,000 for the WBC count.[6] In a large multicenter study of children who presented with temperature greater than 39.5˚C to pediatric clinics and emergency departments, 55 of 60 children with occult bacteremia had a WBC count ≥ 15,000; sensitivity was therefore 91% (55/60)[7] (Table 1).

From the Department of Family Medicine, University of Michigan, Ann Arbor.

Table 1

White Blood Cell Count for Bacteremia in Young Children: Sensitivity and Specificity

	BACTEREMIA	
White blood cell count	*Yes*	*No*
≥15,000	55	275
<15,000	5	177
Total	60	452

Sensitivity = 55/60 = 91%
Specificity = 177/452 = 39%

Data adapted from Bass JW, Steele RW, Wittler RR, et al. Antimicrobial treatment of occult bacteremia: a multicenter cooperative study. *Pediatr Infect Dis J* 1993;12:466-73.

Table 2

Posttest Probability of Bacteremia Among Children Ages 3–36 Months With Fever

	BACTEREMIA		
White blood cell count	*Yes*	*No*	*Total*
≥15,000	36	586	622
<15,000	4	374	378
Total	40	960	1,000

The posttest probability of bacteremia, given a white blood cell count ≥ 15,000 = 36/622 = 6%.

Specificity is defined as the probability of a negative test in those without the disease.[1] In the multicenter study, specificity of a WBC count ≥15,000 was 39%[7] (Table 1).

Sensitivity and specificity can be used to calculate the posttest probability of disease.[1] Suppose that a WBC count on the child in the scenario was 36,000. The pretest probability is 4%; therefore, in a hypothetical group of 1,000 1-year-old children with temperature greater than 39.5˚C, 40 (or 4%) will have occult bacteremia. Of the 40 with occult bacteremia, 91%, or 36, will have a WBC ≥ 15,000, and 39% of the 960 without bacteremia, or 374, will have a WBC count < 15,000 (Table 2). The posttest probability of bacteremia, given a positive WBC count, is known as the positive predictive value and is simply the proportion of those with a positive WBC count who have bacteremia, or 36/622 = 6% (Table 2).

Interval Likelihood Ratios

In contrast to sensitivity and specificity, interval likelihood ratios capture the magnitude of abnormality of test results. Many laboratory tests and imaging studies have continuous or ordinal (ordered categories) outcomes; the magnitude of abnormality on the test is indicative of the magnitude of abnormality of the underlying physiology. When test results with continuous or ordinal outcomes are dichotomized for calculation of sensitivity and specificity, valuable information is lost, because results that are markedly abnormal are lumped together with results that are only mildly abnormal. Interval likelihood ratios, however, assign a specific value to each level of abnormality, and this value can be used to calculate the posttest probability of disease for a given level of a test.

Interval likelihood ratios have widespread applicability because many tests are measured on a continuous or ordinal scale, even those that are often reported as positive or negative. For example, the electrocardiogram stress test is often reported as positive or negative, but the number of millimeters of ST segment depression captures the degree of abnormality of the test. Likelihood ratios can capture this magnitude of abnormality by assigning a specific value to each level of ST segment depression. Table 3 shows tests measured on continuous or ordinal scales but that are often reported as positive or negative.

Table 3

Diagnostic Tests With Continuous or Ordinal Outcomes Typically Reported as Positive or Negative

Test	*Disease or Condition*	*Type of Magnitude*
Straight leg raising	Herniated lumbar disc	Elevation to produce pain, degrees
Pneumatic otoscopy	Otitis media	Movement of tympanic membrane, mm
Anterior drawer	Anterior cruciate tear	Anterior displacement of tibia, mm
Purified protein derivative	Tuberculosis	Diameter of induration, mm
Chest X ray for solitary pulmonary nodule	Lung cancer	Diameter of nodule, cm
Biliary tract ultrasound	Extra-hepatic obstruction	Diameter of biliary tree, cm
Urine culture	Urinary tract infection	Number of bacteria per ml of urine

Table 4

White Blood Cell Count for Bacteremia in Young Children: Likelihood Ratios

WBC count in thousands	Bacteremia (n = 60) #	(%)	No Bacteremia (n = 452) #	(%)	Likelihood Ratio	(95% CI)
<5	0	(0)	12	(3)	0	
5–9.9	0	(0)	87	(19)	0	
10–14.9	5	(8)	78	(17)	.5	(.2, 1.1)
15–19.9	16	(27)	166	(37)	.7	(.5, 1.1)
20–24.9	21	(35)	69	(15)	2.3	(1.5, 3.5)
25–29.9	9	(15)	28	(6)	2.5	(1.2, 5.1)
30–34.9	4	(7)	9	(2)	3.5	(1.1, 10.8)
35–39.9	2	(3)	2	(.4)	7.5	(1.0, 55.1)
>40	3	(5)	1	(.2)	25.0	(2.4, 257.2)

WBC—white blood cell
CI—confidence interval

Data adapted from Bass JW, Steele RW, Wittler RR, et al. Antimicrobial treatment of occult bacteremia: a multicenter cooperative study. *Pediatr Infect Dis J* 1993;12:466-73.

Definition

A likelihood ratio is defined as the probability of a given level of a test result in those with disease divided by the probability of that same result in those without the disease.[1,2] A likelihood ratio indicates how many times more (or less) likely a test result of a given level is obtained in disease than in no disease. When the data for WBC count and bacteremia are presented for each level of WBC, rather than simply as positive or negative, likelihood ratios can be calculated that reflect the magnitude of abnormality. For example, the likelihood ratio for 20,000–24,999 WBCs is the probability of 20,000–24,999 WBCs in those with bacteremia (21/60, or 35%) divided by the probability of 20,000–24,999 WBCs in those without bacteremia (69/452 or 15%), or 35%/15%=2.3 (Table 4). This means that WBC counts between 20,000 and 24,999 are 2.3 times more common in young children with bacteremia than in those without bacteremia. The likelihood ratios increase steadily with increasing WBC count, ranging from 0 for a WBC count <5,000 to 25 for a WBC count >40,000; the magnitude of the likelihood ratio indicates the magnitude of abnormality of the WBC count.

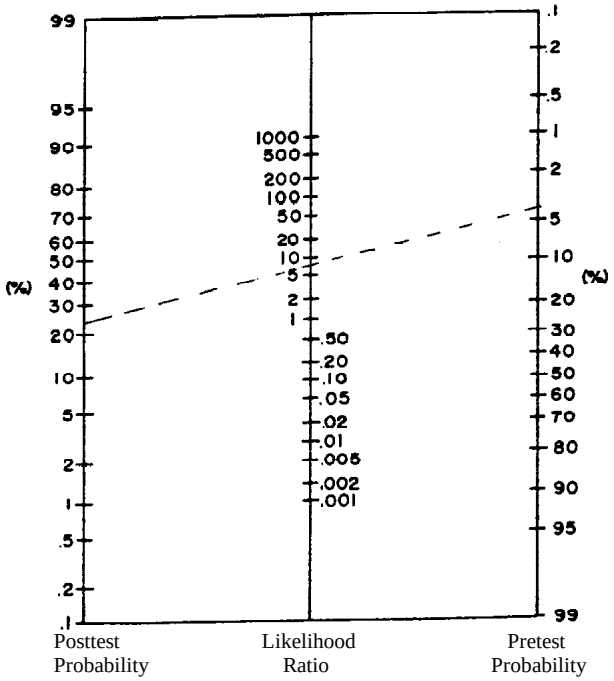
Likelihood ratios can also be used to characterize tests that are dichotomous. The positive likelihood ratio for WBC count, when the results are dichotomized with a cutoff point of 15,000, is defined as the probability of a positive WBC count in those with bacteremia (55/60) divided by the probability of a positive test in those without bacteremia (275/452) or 1.5 (Table 4). This means that a positive WBC count is 1.5 times more common in children with bacteremia than in those without. A positive likelihood ratio can also be calculated as [sensitivity/(1-specificity)]. Since the sensitivity for a WBC count, using the cutoff point of 15,000 is 91%, and the specificity is 39%, the positive likelihood ratio for WBC count is [.91/(1-.39)]=1.5, the same result obtained using the general definition of a likelihood ratio. The negative likelihood ratio for WBC count is defined as the probability of a negative WBC count in those with bacteremia (5/60) divided by the probability of a negative WBC count in those without bacteremia (177/452) or .2 (Table 4). This means that a negative WBC is .2 times as common in children with bacteremia as in children without bacteremia. A negative likelihood ratio can also be calculated as [(1-sensitivity)/specificity].

Using Likelihood Ratios to Determine Probability of Disease

The easiest way to obtain posttest probabilities, given pretest probabilities and likelihood ratios, is to use a nomogram[8] (Figure 1). To use the nomogram, a straight edge is anchored at the right border of the nomogram, which is the estimated pretest probability. Then, the straight edge is rotated to intersect the likelihood ratio,

Figure 1

A Nomogram for Using Likelihood Ratios to Obtain Posttest Probabilities of Disease



The straight line connects a pretest probability of 4% (on the right axis) to the likelihood ratio of 7.5 (on the middle axis), giving a posttest probability of 24% (on the left axis).

in the middle of the nomogram. The posttest probability is the point of intersection of the straight edge and the left edge of the nomogram.

What is the posttest probability of bacteremia, given a pretest probability of 4% and a WBC count of 36,000? The likelihood ratio for a WBC count of 36,000 is 7.5 (Table 4). The posttest probability of bacteremia, given a pretest probability of 4% and a WBC count with a likelihood ratio of 7.5, is 24%. This result is dramatically different from the 6% posttest probability that was calculated using sensitivity and specificity, and this difference has important clinical implications. Empiric intramuscular antibiotics to prevent meningitis may be justifiable at a probability of bacteremia of 24% but may not be justifiable at a probability of 6%; the potential benefit of preventing meningitis may outweigh the pain and risk of side effects when the probability of bacteremia is high but not when the probability is low.

Interval likelihood ratios are thus more informative than sensitivity and specificity, because the posttest probabilities of disease calculated from them cover a wider range than posttest probabilities calculated from sensitivity and specificity. Since clinical decisions are based on estimates of the probability of disease, decisions derived from the use of interval likelihood ratios may be different from decisions derived from test results characterized by sensitivity and specificity.

A nomogram may not always be available to determine posttest probabilities. The posttest probability of disease can also be calculated from the likelihood ratio version of Bayes theorem:

(Pretest odds) x (likelihood ratio)=posttest odds

To use this formula, pretest probability must be converted to pretest odds, and posttest odds must be converted back to posttest probability. Odds are a mathematical way of expressing the probability of an event divided by the probability of the event not happening:

Odds=probability/(1-probability)

The pretest probability of bacteremia is 4% (or .04). The pretest odds are then .04/.96. Using the likelihood ratio of 7.5 for a WBC count of 36,000, the posttest odds are then: (.04/.96) x 7.5=.31. By rearranging the formula for odds, it can be shown that:

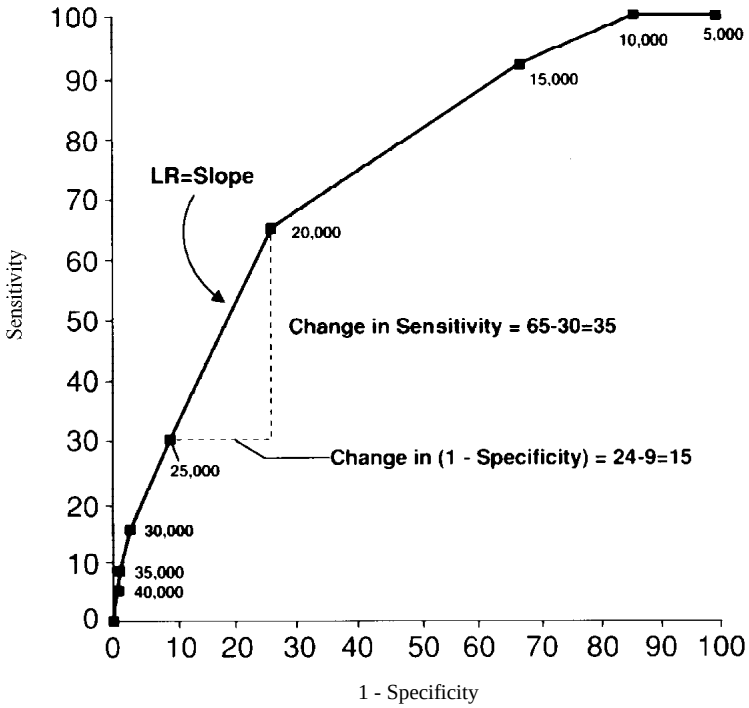
Probability=odds/(odds+1)

The posttest probability of bacteremia is therefore .31/(.31+1)=.24, or 24%, the same result obtained by using the nomogram.

The likelihood ratio version of Bayes theorem (equation 1) has the virtue of demonstrating two important conclusions regarding diagnostic tests. First, the posttest probability depends on the test characteristics and the pretest probability. For any given likelihood ratio, the higher the pretest probability, the higher the posttest probability and vice versa. Second, test results with likelihood ratios close to 1.0 provide little diagnostic information, because the posttest probability is unchanged, compared with the pretest probability.

Figure 2

ROC Curve for WBC Count for Diagnosis of Bacteremia



ROC—receiver operating characteristic
WBC—white blood cell

Each point on the ROC curve marks the sensitivity and (1-specificity) corresponding to each WBC count cutoff point. The interval likelihood ratio for the 20,000–24,999 WBC count category is the slope of the ROC curve between the 20,000 and the 25,000 cutoff points. The slope is calculated as the change in the y axis (sensitivity) divided by the change in the x axis (1-specificity) for the two cutoff points or 35/15=2.3.

Receiver Operating Characteristic Curves

Receiver operating characteristic (ROC) curves are a graphic method of showing the characteristics of diagnostic tests with continuous or ordinal results. An ROC curve is a graph of the true positive rate (sensitivity) and the false positive rate (1-specificity) at each possible cutoff point of the test results.

Figure 2 shows the ROC curve for WBC count for diagnosis of bacteremia in young children. Each point on the curve corresponds to a different cutoff point. The slope of the curve between two contiguous cutoff points equals the interval likelihood ratio for that interval. For example, the likelihood ratio for the interval 20,000–24,999 WBCs is calculated as the change in the y axis (sensitivity) corresponding to 20,000 as a cutoff point (65%) and 25,000 as a cutoff point (30%) divided by the change in the x axis (1-specificity) corresponding to those cutoff points or 76% and 91%, respectively. The interval likelihood ratio is then (65%-30%)/(91%-76%)=35/15=2.3, the same result shown in Table 4.

The closer an ROC curve is to the upper left corner of the graph (the greater the area under the curve), the better the diagnostic test, since there will be smaller decreases in sensitivity as specificity increases and vice versa. In comparing two or more diagnostic tests, the test with the greatest area under its ROC curve is a better test.

Research Articles on Diagnostic Tests

Research articles on diagnostic tests should report data to facilitate calculation of likelihood ratios. For tests with ordinal outcomes, the number of patients with and without disease should be reported at each level of the test, and likelihood ratios should be reported for each level.

For tests with continuous outcomes, such as blood counts or serum chemistry values, the data should be grouped into multiple categories, and the number of patients with and without disease should be reported in each category. How many categories should be created? How many subjects should be in each category? There are no infallible rules but two basic approaches. The goal is to identify homogenous groups for which there are important differences in the probability of the outcome between the groups but not within them.[9]

The first approach is to identify categories that correspond to natural breaks in the data. This method has the advantage of being driven by the raw data, not theory. The disadvantage is that the results may not be generalizable if the data set has idiosyncrasies.

The second approach is to create clinically meaningful categories, based on prior knowledge. The disadvantage of this method is that important new findings may not be detected, if existing knowledge is limited.

One method that should not be used is categorizing the data by percentiles, eg, quartiles or quintiles. Categories developed by percentiles make it difficult to compare likelihood ratios across different studies of the same diagnostic test, because each study will have a different proportion of subjects at any given level of the diagnostic test.

One should choose a number of categories large enough to provide a wide range in the likelihood ratios but small enough to ensure stability of the likelihood ratios in each category. A small number of categories will mask the virtues of likelihood ratios because the high and low likelihood ratios corresponding to the data at the extremes of the distribution will be "diluted" by the likelihood ratios in the middle of the distribution. If too many categories are chosen, there will be few subjects in each category, and the estimates of the likelihood ratios will be unstable. Collapsing contiguous categories at the extremes of the distribution, where there are likely to be few subjects, will often provide more stable estimates of likelihood ratios at these values.

If a test is intrinsically dichotomous, the results may be reported in terms of either sensitivity and specificity, or positive and negative likelihood ratios, because there are no important advantages of likelihood ratios in this case.

Limitations

There are several important limitations to the use of likelihood ratios to calculate posttest probabilities of disease in primary care populations. The first limitation is that most of the existing data on pretest probabilities were obtained from tertiary care populations and may not be generalizable to primary care populations.[10,11] For example, the pretest probability of bacteremia of 4% used in the clinical scenario in this article was obtained from a sample of patients in emergency departments and pediatric clinics and may not be applicable to children with fever in family practice office settings. This highlights the need for research on pretest probabilities in primary care settings.

Second, although test results at the extremes of the distribution provide the greatest diagnostic information, because they have the most extreme likelihood ratios, the estimates of likelihood ratios at these extreme values are very imprecise (ie, they have wide confidence intervals) due to the sparse data at the extremes.[12,13] Although the point estimate for the likelihood ratio for >40,000 WBCs is 25.0, the 95% confidence interval is 2.4 to 257.2 (Table 4). The data in Table 4 are presented as they were shown in the original research report, but more stable estimates could be obtained by collapsing the bottom two and the upper three categories, yielding a total of six, rather than nine, categories.

Third, although test results are most helpful when they are extreme, one is much more likely to obtain a value in the middle of the distribution, where the likelihood ratios are closer to 1 and, therefore, less useful in changing the estimate of the probability of disease. The probability of obtaining a WBC count of 35,000–39,999 (likelihood ratio of 7.5) is 4/512 or .7%, but the probability of obtaining a WBC count of 15,000–19,999 (likelihood ratio of .7) is 182/512 or 36%.

Finally, likelihood ratios may change with pretest probability of the disease. It can be shown mathematically that likelihood ratios are independent of pretest probability. However, in actual practice, likelihood ratios may vary with pretest probability, due to differences in the spectrum of disease in population subgroups, such as between men and women.[14] Further, due to the diagnostic misclassification that is introduced when a continuous outcome is changed into discrete categories, likelihood ratios can vary enormously with differences in pretest probability, even in the absence of a change in the spectrum of disease.[15] This is an important limitation, since it means that one cannot assume a constant likelihood ratio for a test across population subgroups. This limitation applies to sensitivity and specificity as well and is not unique to likelihood ratios.

Conclusions

Interval likelihood ratios capture the magnitude of abnormality of a diagnostic test. Posttest probabilities derived from likelihood ratios can be substantially different from probabilities derived from sensitivity and specificity; these differences may have important clinical ramifications. Researchers should report data on diagnostic tests to facilitate calculation of likelihood ratios, and clinicians should use likelihood ratios, whenever possible, to make clinical decisions involving diagnostic tests.

Acknowledgments: A lecture based on information in this manuscript has been presented at 1) the University of Michigan Department of Family Medicine grand rounds in February 1996, 2) the University of Michigan Outpatient Attending Lecture Series, several times, and 3) the 1997 12th Annual Primary Care Research Methods and Statistics Conference in San Antonio, Tex.

Correspondence: Address correspondence to Dr Sonis, University of Michigan, Department of Family Medicine, 1018 Fuller Street, Ann Arbor, MI 48109-0708. 313-998-7120. Fax: 313-998-7335. E-mail: jsonis@umich.edu.

References

1. Fletcher RH, Fletcher SW, Wagner EH. Clinical epidemiology: the essentials, third edition. Philadelphia: Williams and Wilkins, 1996:64-7.
2. Jaeschke R, Guyatt GH, Sackett, DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results, and will they help me in caring for my patients? JAMA 1994;271:703-7.
3. Giard RW, Hermans J. The diagnostic information of tests for the detection of cancer: the usefulness of the likelihood ratio concept. Eur J Cancer 1996;32A:2042-8.
4. Baraff LJ, Bass JW, Fleisher GR, et al. Practice guideline for the management of infants and children 0 to 36 months of age with fever without source. Ann Emerg Med 1993;22:1198-1220.
5. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. N Engl J Med 1980;302:1107-9.
6. Kramer MS, Tange SM, Mills EL, Ciampi A, Bernstein ML, Drummond KN. Role of the complete blood count in detecting occult fecal bacterial infection in the young febrile child. J Clin Epidemiol 1993;46:349-57.
7. Bass JW, Steele RW, Wittler RR, et al. Antimicrobial treatment of occult bacteremia: a multicenter cooperative study. Pediatr Infect Dis J 1993;12:466-73.
8. Fagan TJ. Nomogram for Bayes's theorem. (Letter) N Engl J Med 1975;293:257.
9. Rothman KJ, Greenland S. Modern epidemiology, second edition. Philadelphia: Lippincott-Raven, 1998.
10. Dans AL, Dans LF, Guyatt GH, Richardson S. Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. Evidence-based Medicine Working Group. JAMA 1998;279:545-9.
11. Sox HC Jr, Hickam DH, Marton KI, et al. Using the patient's history to estimate the probability of coronary artery disease: a comparison of primary care and referral. Am J Med 1990;89:7-14.
12. Dujardin B, Van den Ende J, Van Gompel A, Unger J-P, Van der Stuyft P. Likelihood ratios: a real improvement for clinical decision making? Eur J Epidemiol 1994;10:29-36.
13. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. J Clin Epidemiol 1991;44:763-70.
14. Moons KG, van Es GA, Deckers JW, Habbema JD, Grobbee DE. Limitations of sensitivity, specificity, likelihood ratio, and Bayes' theorem in assessing diagnostic probabilities: a clinical example. Epidemiology 1997;8:12-7.
15. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios, and predictive values with disease prevalence. Stat Med 1997;16:981-91.