# Likelihood Ratio: A Powerful Tool for Incorporating the Results of a Diagnostic Test Into Clinical Decisionmaking

*From the Department of Emergency Medicine, University of California San Diego Medical Center, San Diego, CA,[*] and the Department of Emergency Medicine, Butterworth Hospital/Michigan State University, Grand Rapids, MI.[‡]*

**Stephen R Hayden, MD[*]**
**Michael D Brown, MD[‡]**

*Address for reprints: Stephen R Hayden, MD, Department of Emergency Medicine, UCSD Medical Center, Mail Code 8676, 200 West Arbor Drive, San Diego, CA 92103.*

**See related article, p. 565.**

[Hayden SR, Brown MD: Likelihood ratio: A powerful tool for incorporating the results of a diagnostic test into clinical decisionmaking. *Ann Emerg Med* May 1999;33:575-580.]

Diagnostic tests rarely confirm or exclude the presence of disease with certainty. More frequently, the results of a diagnostic test are used by clinicians to strengthen their estimate that a disease is likely or unlikely in a particular patient. Diagnostic test results may accomplish this by virtue of being labeled "positive," "negative," "high probability," or "low probability." Such labels do not, however, guarantee anything about the magnitude by which the test result in question strengthens the clinical assessment. Although clinicians are accustomed to the terms "sensitivity" and "specificity" as measures of a diagnostic test's reliability, these characteristics of a test have certain disadvantages. These terms pertain to the likelihood of a particular test result in patients independently known to have or not to have the disease in question. They do not in themselves tell the clinician how likely the individual patient is to have the disease as a result of that test result. Furthermore, when a diagnostic test has more than 2 possible results, such as in the case of a ventilation-perfusion scan for pulmonary embolus, it is no longer possible to define results as simply "positive" or "negative." Under such circumstances, the terms "sensitivity" and "specificity" are no longer applicable.

The magnitude of change from a clinician's initial (pretest) assessment of the probability of disease to the likelihood of disease after knowing the result of a diagnostic test (posttest probability) is represented by the *likelihood ratio* (LR). The LR can be derived from the test's sensitivity and specificity[1] or from the primary data as reported in studies on diagnostic test accuracy. A diagnostic test is useful only if the result substan-

tially alters the pretest probability. This is determined by the accuracy with which it identifies the disease of interest.

How will the result of a diagnostic test influence clinical decisionmaking? First, it is necessary to estimate the pretest probability. In caring for individual patients, before performing a test, the clinician has a level of disease probability in mind, which requires no further testing and prompts treatment. This is commonly referred to as the *"treatment" threshold*. Similarly, a clinician has a level of disease probability in mind that effectively rules out the disease and requires no further testing. This is termed the *"test" threshold* (Figure 1).

In most cases, clinical thinking revolves around qualitative assessments of "high," "intermediate," and "low" probability of disease. As will be shown, LRs can be used to demonstrate that conclusions made on the basis of diagnostic test results tend to hold up over a broad range of pretest and posttest probability values. LRs can also be used to delineate the minimum performance requirements for a diagnostic test when more precisely defined decision thresholds are required.

In some situations, the determination of pretest probabilities may be assisted by published data or by institution-specific data. This is particularly important when LRs derived from studies on the diagnostic accuracy of elements of the clinical evaluation are to be used in patient care and decisionmaking.

## WHAT ARE LIKELIHOOD RATIOS?

The LR is a powerful measure of the accuracy of a diagnostic test. It is the ratio of the probability of a given test result in patients *with disease* to the probability of the same test result in patients *without disease.* The LR for a given test result indicates how much that result will raise or lower the probability of disease. It can be used with a nomogram developed by Fagan[2] (Figure 2) to establish the posttest probability corresponding to any pretest probability and for any test result.

## HOW ARE LIKELIHOOD RATIOS DERIVED FROM PUBLISHED DATA?

LRs are not always directly reported in studies of the accuracy of diagnostic tests. It is, however, usually possible for the reader to calculate them using data commonly reported in such studies.

Authors report the results of studies in a variety of ways. When only 2 test results are possible, they are commonly labeled "positive" versus "negative." In such cases, the results of studies on accuracy are frequently reported as "sensitivity," "specificity," "positive predictive value," and "negative predictive value." This terminology and method of reporting represents a holdover from the realm of "diagnostics" originally developed in the context of assays for presence or absence of substances in chemistry research. It is much less appropriate when the results of diagnostic tests are to be considered for use in clinical decisionmaking. What a clinician needs to know is how frequently a particular diagnostic test result occurs in patients with a particular disease, condition, or injury compared with its frequency in patients without that clinical entity.

The following examples illustrate how LRs can be derived from data as reported in published studies on diagnostic tests.

### Using a 2×2 table

When only 2 possible test results are under consideration, a 2×2 table allows rapid calculation of the LRs (Table 1). An example of this is a "rapid strep" test for presence or absence of group A β-hemolytic *Streptococcus* (GABHS) in the oropharynx of a patient with sore throat.

**Figure 1.**

*Decisionmaking thresholds. The numbers assigned to test and treatment thresholds will vary from one clinical setting to another (ie, 1%, 5%, 10%, and so on for test threshold, 80%, 90%, 99%, and so on for treatment threshold). They may differ for different diseases, and will be influenced by the clinical experience of the practitioner and by patient preferences. These thresholds should be determined by the clinician before obtaining the diagnostic test result. If no test result results in a posttest probability that either mandates treatment, consideration of alternative diagnoses, or discharge home, then the test will not aid clinical decisionmaking.*
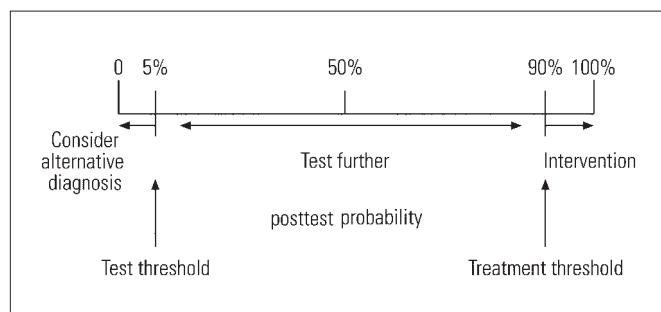
When study data are reported as "sensitivity" and "specificity," the LRs can be derived from this data as follows:

$$LR\,[+] = \frac{Sensitivity}{1-Specificity}$$

$$LR\,[-] = \frac{1-Sensitivity}{Specificity}$$

where LR[+] is the LR of a positive test result, and LR[–] is the LR of a negative test result.

As is true of the sensitivity and specificity of a diagnostic test, LRs are stable to changes in prevalence of disease as opposed to "positive predictive value" and "negative predictive value," which are dependent on the prevalence of disease in a population.

**Other LR calculations**

LRs can also be calculated when more than 2 test results are possible. An example of this is the conventional reporting of results of ventilation-perfusion scans in the evaluation of patients with suspected pulmonary embolism (Table 2). In this case, the terms "sensitivity" and "specificity" are not applicable. Similarly, LRs can be calculated for test results that are grouped into specific intervals, assuming there is no overlap of the intervals (ie, test result of 0-5, 6-10, 11-15, and so on.)

Finally, LRs also can be calculated for diagnostic test results when the results are continuous variables. An example of this is the peripheral WBC count in patients with suspected acute appendicitis or in febrile infants at risk of occult bacteremia. The valid approach to deriving LRs in this situation is more involved, is not always possible from the data in published studies, and is beyond the scope of this discussion. A recent article illuminates why

**Figure 2.**
*Nomogram for interpreting diagnostic test results.*

**Table 1.**
*Deriving LRs when only 2 test results are possible.*

| Group | Disease+ | Disease– | Total |
|---|---|---|---|
| Test result positive | a | b | a+b |
| Test result negative | c | d | c+d |
| **Total** | a+c | b+d | n[*] |

**a, b, c,** and **d** represent the number of study patients in each of the categories of the 2×2 table. "**Disease+**" and "**Disease–**" refer to the categorization of the individual study patients by the criterion standard used in the study.
Likelihood ratio positive test: (LR[+])=(a/a+c)/(b/b+d).
Likelihood ratio negative test (LR[–])=(c/a+c)/(d/b+d).
[*]n=a+b+c+d.

LRs derived from a simple dichotomization of such results, such as defining WBC counts above some value as "positive" and those below it as "negative," are misleading and tend to exaggerate the diagnostic accuracy of such tests.[3]

## HOW ARE LIKELIHOOD RATIOS APPLIED IN CLINICAL DECISIONMAKING?

A diagnostic test that is useful would ideally have either a high LR or a low LR. As the LR approaches 1.0, the utility of the test decreases to zero. Once the LR is known for a given diagnostic test result, use of the nomogram developed by Fagan[2] is the simplest method for the clinician to arrive at a posttest probability. Use of the nomogram avoids the necessity of an extra set of computations and streamlines the process of using LRs. In most cases, practitioners formulate their clinical assessments of individual patients, not in terms of exact numbers for pretest probabilities, but as qualitative categories corresponding to "low probability," "intermediate probability," and "high probability" of disease. Similarly, "test" and "treatment" thresholds are also largely perceived in qualitative terms. In some cases, more precise posttest thresholds can be defined. For example, the threshold for taking a patient with suspected appendicitis to the operating room is typically 90%, and a negative laparotomy rate no greater than 10% is acceptable to most surgery departments. A decision analysis on the cost-effectiveness of admissions to the hospital of patients with chest pain concluded that the rate for missed myocardial infarction

should not be higher than 2% in patients discharged home without admission.[4]

Table 3 provides a guide to determining the effect of LRs of different magnitude on the posttest probability of disease. Table 4 demonstrates that a test result with a LR of 10 converts qualitative pretest clinical assessments of "low" to "intermediate" as defined over a broad range of values, to a posttest likelihood of disease in the "intermediate" to "high" probability range. LRs of greater than 10 will have an even more dramatic effect. Similarly, a test result with a LR of 0.1 will reduce even a very high (up to 90%) level of pretest suspicion to a level of 50% or lower. This same principle is illustrated in a different way in Figures 3 and 4 using the Fagan nomogram introduced previously.

## EXAMPLE

Consider the case of an 8-year-old boy who presents to the ED complaining only of sore throat and fever. On examination he has an erythematous pharynx with mild anterior cervical adenopathy. Because physicians may overestimate GABHS infection,[5,6] you have adopted a strategy that uses diagnostic testing as opposed to empiric treatment for pediatric patients with sore throat.[7,8] You are able to arrange good follow-

**Table 2.**
*Calculating LRs when more than 2 test results are possible.*

| Probability Category | Disease+ | Disease– | Total |
|---|---|---|---|
| High | a | b | a+b |
| Intermediate | c | d | c+d |
| Low | e | f | e+f |
| **Total** | x | y | n* |

**a, b, c, d, e,** and **f** represent the number of patients within each category in the table. The categories of high, intermediate, or low probability could also be discrete intervals of test results (0-5, 6-10, 11-15, etc). "**Disease+**" and "**Disease–**" are determined by the criterion standard used in the study.
LR for high probability result=(a/x)/(b/y). LR is likelihood of a high probability test result when disease *is* present divided by likelihood of a high probability test result when *no* disease is present.
LR for intermediate probability result=(c/x)/(d/y). LR is likelihood of an intermediate probability test result when disease *is* present divided by likelihood of an intermediate probability test result when *no* disease is present.
LR for low probability result=(e/x)/(f/y). LR is the likelihood low probability test result when disease *is* present divided by the likelihood of a low probability test result when *no* disease is present.
*n=a+b+c+d+e+f.

**Table 3.**
*The impact of likelihood ratios of different magnitude on the posttest probability of disease.*

| High LRs | Low LRs | Effect on Posttest Probability |
|---|---|---|
| >10 | <.1 | Large |
| 5–1 | .1–.2 | Moderate |
| 2–5 | .2–.5 | Small |
| 1 | 1 | No change |

**Table 4.**
*Effect of LRs of 10 and .1 on qualitative ranges of pretest probability.*

| LR | Pretest Probability % | Posttest Probability % |
|---|---|---|
| 10 | 10–30 (low) | 53–80 (moderate to high) |
| 10 | 30–60 (intermediate) | 80–95 (high) |
| .1 | 30–60 (intermediate) | 3–12 (low) |
| .1 | 60–90 (high) | 12–50 (low to intermediate) |

up for your ED population, and your clinical estimate of the likelihood of a positive result on standard throat culture for GABHS in this patient is about 50%. Using published data for the second-generation enzyme immune assay used in your ED,[9] you calculate a LR of 20 for a positive result, which is in the range having a large effect on posttest probability. However, the LR for negative strep screen is .2, which is only in the moder-

ately useful range. Will test results with these LRs affect your decisionmaking for this child?

We invite readers, using the nomogram in Figure 2, to draw a line connecting the clinical estimate of a 50% chance of GABHS on the left-hand column (estimated pretest probability) through the LRs of 20 and .2, respectively. These lines intersect the right-hand column and give the posttest probability of a positive culture result for

**Figure 3.**
*A LR of .1 will convert a pretest probability in the 30% to 90% range (intermediate to high) to a posttest probability in the 13% to 50% range (low to intermediate).*
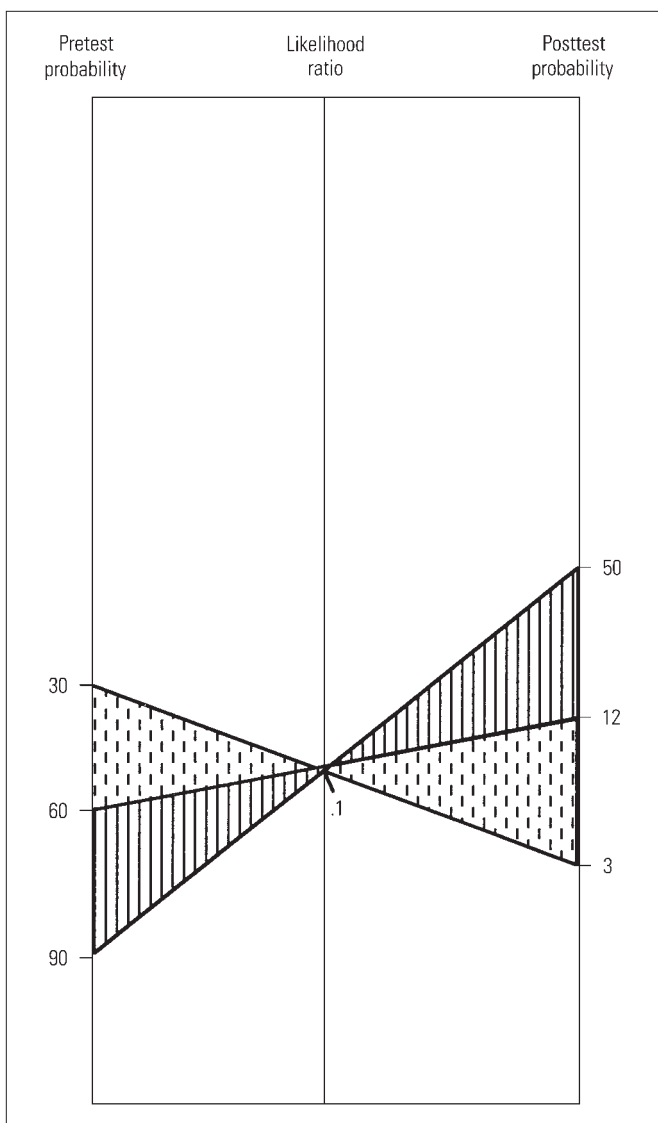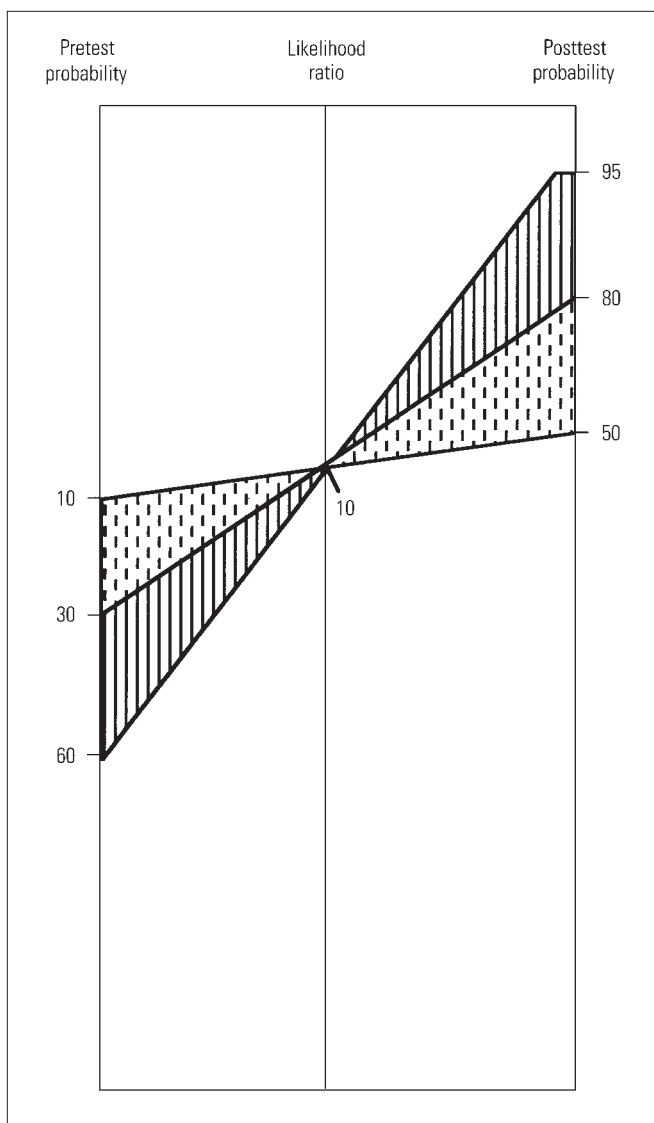


**Figure 4.**
*A LR of 10 will convert a pretest probability in the 10% to 60% range (low to intermediate) to a posttest probability in the 50% to 95% range (moderate to high).*

GABHS that correspond to a positive and negative rapid strep test result.

Readers will find that, if the 8-year-old patient has a positive rapid strep test result, he has a 97% chance of positive growth of GABHS on a standard culture. Most practitioners, given this result, would feel very confident treating the child with no further testing. On the other hand, a negative rapid strep test result would still leave the patient with a 20% chance of having a GABHS-positive culture. If readers were uncomfortable with the idea of not treating 1 of 5 cases of GABHS, a standard throat culture would then be indicated.

As a further exercise, readers may want to experiment, using the nomogram, to determine how low the clinical estimate would have to be for a LR of 20 to still result in an unequivocal decision to treat. To do this, readers must decide what an appropriate "treatment threshold" is.

Similarly, given a negative rapid strep test result, how low would the clinical estimate have to be for a LR of .2 to result in a decision to simply send the patient home with no further testing and no treatment? If that threshold clinical estimate were already "low" (ie, <30%), it might well be decided that the rapid strep test has a very straightforward impact on practice. There being no clinical setting in which the test would lead you to a "no treatment, no testing" option, you will treat the patient when the rapid test result is positive and test (standard throat culture) when it is not. The issue of whether to initiate treatment pending culture results is an independent question, subject to a differently directed literature review.

In summary, the use of likelihood ratios for clinical decisionmaking in the care of individual patients requires that practitioners formulate qualitative estimates of disease likelihood regarding patients for whom they are considering diagnostic testing beyond the level of primary clinical evaluation. This requirement is an included feature of accepted standards of mature clinical practice. LRs enable a clearer and more precise evaluation of the appropriate impact of diagnostic test results on clinical decisionmaking.

## REFERENCES

1. Jaeschke R, Guyatt GH, Sackett DL, et al: User's guides to the medical literature, III: How to use an article about a diagnostic test. *JAMA* 1994;271:703-707.

2. Fagan TJ: Nomogram for Baye's theorem (C). *N Engl J Med* 1975;293:257.

3. Tandberg D, Deely JJ, O'Malley AJ: Generalized likelihood ratios for quantitative diagnostic test scores. *Am J Emerg Med* 1997;15:694-699.

4. Wears RL, Li S, Hernandez JD, et al: How many myocardial infarctions should we rule out? *Ann Emerg Med* 1989;18:953-963.

5. Poses RM, Cebul RD, Collins M, et al: The accuracy of experienced physicians' probability estimates for patients with sore throats. *JAMA* 1985;254:925-929.

6. Putto A: Febrile exudative tonsillitis: Viral or streptococcal. *Pediatrics* 1987;80:6-11.

7. Pichichero ME: Group A streptococcal tonsillopharyngitis: Cost-effective diagnosis and treatment. *Ann Emerg Med* 1995;25:390-403.

8. Bisno AL, Gerber MA, Gwaitney JM Jr, et al: Diagnosis and management of group A streptococcal pharyngitis: A practice guideline. *Clin Infect Dis* 1997;25:574-583.

9. Roe M, Kishiyama C, Davidson K, et al: Comparison of BioStar Strep A OIA optical immune assay, Abbot TestPack Plus StrepA, and culture with selective media for diagnosis of group A streptococcal pharyngitis. *J Clin Microbiol* 1995;33:1551-1553.