

BRIEF REPORTS

Simplifying Likelihood Ratios

Steven McGee, MD

Likelihood ratios are one of the best measures of diagnostic accuracy, although they are seldom used, because interpreting them requires a calculator to convert back and forth between "probability" and "odds" of disease. This article describes a simpler method of interpreting likelihood ratios, one that avoids calculators, nomograms, and conversions to "odds" of disease. Several examples illustrate how the clinician can use this method to refine diagnostic decisions at the bedside.

KEY WORDS: likelihood ratio; diagnostic accuracy.

J GEN INTERN MED 2002;17:647-650.

Likelihood ratios (LRs) constitute one of the best ways to measure and express diagnostic accuracy. Despite their many advantages, however, LRs are rarely used, primarily because interpreting them requires a calculator to convert back and forth between probability of disease (a term familiar to all clinicians) and odds of disease (a term mysterious to most people other than statisticians and epidemiologists). Although nomograms can circumvent these calculations,^{1,2} these nomograms are rarely accessible at the bedside and are seldom used. This article describes a simpler method of interpreting LRs, one that avoids calculators, nomograms, and conversions to odds of disease, and one that illustrates how LRs can refine diagnostic decisions at the bedside.

DEFINITION OF LR

The LR of any clinical finding is the probability of that finding in patients with disease divided by the probability of the same finding in patients without disease:

$$LR = \frac{\text{probability of finding in patients with disease}}{\text{probability of same finding in patients without disease}}$$

For example, among patients with abdominal distension who undergo ultrasonography, the physical sign "bulging flanks" is present in 80% of patients with confirmed ascites and in 40% without ascites (i.e., the distension is from fat or gas). The LR for "bulging flanks" in

detecting ascites, therefore, is 2.0 (i.e., 80% divided by 40%). Similarly, if the finding of "flank tympany" is present in 10% of patients with ascites but in 30% with distension from other causes, the LR for "flank tympany" in detecting ascites is 0.3 (i.e., 10% divided by 30%).

LRs may range from 0 to infinity. Findings with LRs greater than 1 argue for the diagnosis of interest; the bigger the number, the more convincingly the finding suggests that disease. Findings whose LRs lie between 0 and 1 argue against the diagnosis of interest; the closer the LR is to 0, the less likely the disease. Findings whose LRs equal 1 lack diagnostic value.

CONVENTIONAL APPLICATION OF LRS

How much does the finding of bulging flanks (LR = 2.0) argue for ascites, and how much does the finding of flank tympany (LR = 0.3) argue against it? To answer these questions using traditional methods, clinicians must first identify the pretest probability (or prevalence) of ascites in their practice and then perform 3 calculations. For example, if about 2 out of every 5 patients with abdominal distension have ascites, the pretest probability is 40%. The traditional way of applying the finding of bulging flanks (LR = 2.0) is to then convert pretest probability (P_{pre}) to pretest odds (O_{pre}), using $O_{pre} = P_{pre}/(1 - P_{pre})$, then multiply the pretest odds (O_{pre}) by the LR for the finding to derive the posttest odds (i.e., $O_{post} = LR \times O_{pre}$), and then convert posttest odds back to posttest probability, using $P_{post} = O_{post}/(1 + O_{post})$. Using a calculator to complete these 3 calculations, we discover that the finding of bulging flanks (LR = 2.0) increases the probability of ascites from 40% to 57% (i.e., pretest odds = $0.4/(1 - 0.4) = 0.667$; posttest odds = $0.667 \times 2.0 = 1.333$; posttest probability = $1.333/(1 + 1.333) = 0.57$ or 57%). Similar calculations reveal that the finding of flank tympany decreases the probability of ascites from 40% to 17%.

A SIMPLER METHOD OF APPLYING LRS

A simpler method avoids these calculations by using the estimates shown in Table 1. According to these estimates, which are independent of pretest probability, a finding with an LR of 2.0 increases the probability of disease about 15%, and a finding with an LR of 0.3 decreases probability of disease about 25%. Therefore, bulging flanks increases the probability of ascites from 40%

Received from the University of Washington Seattle-Puget Sound VA Health Care System, Seattle, Wash.

Address correspondence and requests for reprints to Dr. McGee: University of Washington Seattle-Puget Sound VA Health Care System (S-111 GIMC), 1660 S. Columbian Wy., Seattle, WA 98108 (e-mail: Steven.McGee@med.va.gov).

to about 55% (i.e., $40 + 15 = 55\%$, only 2% lower than the calculated answer), and flank tympany decreases it from 40% to about 15% (i.e., $40 - 25 = 15\%$, only 2% lower than the calculated answer). As long as the clinician rounds estimates of posttest probability more than 100% to an even 100% and those of less than 0% to an even 0%, these estimates are accurate to within 10% of the calculated answer for all pretest probabilities between 10% and 90%. The average error is only 4%.

Table 1 is easy to recall at the bedside by simply remembering 3 specific LRs—2, 5, and 10—and the first 3 multiples of 15 (i.e., 15, 30, and 45). An LR of 2 increases probability 15%, one of 5 increases it 30%, and one of 10 increases it 45%. For those LRs between 0 and 1, the clinician simply inverts 2, 5, and 10 (i.e., $1/2 = 0.5$, $1/5 = 0.2$, $1/10 = 0.1$). Just as the LR of 2.0 increases probability 15%, its inverse, 0.5, decreases probability 15%. Similarly, an LR of 0.2 (the inverse of 5) decreases probability 30%, and a LR of 0.1 (the inverse of 10) decreases it 45%. These benchmark LRs can be used to approximate the remainder of Table 1.

Although this method is inaccurate for pretest probabilities less than 10% or greater than 90%, this is not a disadvantage, because these polar extremes of probability indicate diagnostic certainty for most clinical problems, making it unnecessary to order further tests (and apply additional LRs).

DERIVATION OF TABLE 1

The estimates shown in Table 1 agree with the calculated answers despite ignoring pretest probability and odds of disease because of the peculiar S-shaped

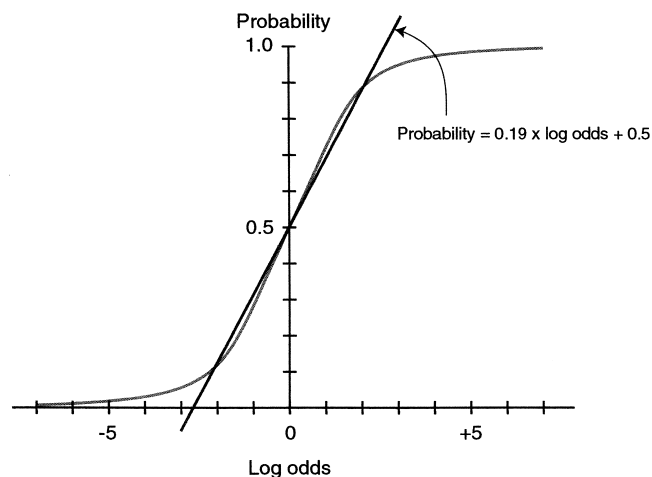


FIGURE 1. The S-shaped curve describes the actual relationship between probability and log odds, and the straight line is the estimate of the nearly linear portion of this curve between probabilities of 10% and 90%. The S-shaped curve is the logistic function $P = 1/(1 + e^{-z})$, where $z = \log \text{odds}$.

relationship between probability and log odds* (Fig. 1), which, between probabilities of 10% and 90%, is nearly linear. Recalling the equation $O_{\text{post}} = \text{LR} \times O_{\text{pre}}$ and its logarithmic equivalent ($\log O_{\text{post}} = \log \text{LR} + \log O_{\text{pre}}$), simple substitution of the linear approximation from Figure 1 reveals:

$$\text{Change in probability} \approx 0.19 \times \log \text{LR}.$$

In other words, regardless of the patient's pretest probability, the change in probability from a finding is approximated by a constant ($0.19 \times \log \text{LR}$). These estimates, rounded off to the nearest 5% for easy recall, appear in Table 1.

USING THIS METHOD IN DAILY PRACTICE

The main advantage of LRs (over other measures of diagnostic accuracy, such as sensitivity and specificity) is that clinicians can use them to quickly compare different diagnostic strategies and thus refine clinical judgment. Several types of these comparisons appear in Table 2 and are described further below.

The most common type of comparison examines different tests for the same diagnosis. For example, Table 2 displays the LRs for different physical signs in detecting ascites. "Positive LR" describes how probability of disease shifts when the finding is present; "negative LR" describes how probability of disease shifts when it is absent. With only a glance at this Table, clinicians can determine that among these 5 physical signs the most compelling argument for ascites (because it has the greatest LR) is the positive fluid wave, which increases the

Table 1. Likelihood Ratios and Bedside Estimates

Likelihood Ratio	Approximate Change in Probability (%)*
Values between 0 and 1 decrease the probability of disease	
0.1	-45
0.2	-30
0.3	-25
0.4	-20
0.5	-15
1	0
Values greater than 1 increase the probability of disease	
2	+15
3	+20
4	+25
5	+30
6	+35
7	
8	+40
9	
10	+45

The text describes how to easily recall these estimates.

*"log" refers to the natural logarithm, i.e., \log_e .

Table 2. Types of Comparisons Using Likelihood Ratios

Finding	Positive LR	Negative LR
Compare the accuracy of different diagnostic tests for the same diagnosis		
Physical findings, detecting ascites ^{3,4}		
Bulging flanks	2.0	0.3
Flank dullness	2.0	0.3
Edema	3.8	0.2
Fluid wave	6.0	0.4
Shifting dullness	2.7	0.3
Compare the diagnostic impact of a test's positive result to its negative result		
Hyperresonance, detecting chronic airflow obstruction ⁵	5.1	NS
Pleuritic component to chest pain, detecting myocardial Infarction ⁶	0.2	1.3
Compare the accuracy of the same test for different definitions of disease		
Late-peaking systolic murmur ⁷		
Detecting severe aortic stenosis	3.0	0.2
Detecting moderate-to-severe aortic stenosis	24.2	0.3
Compare the accuracy of different levels of the same test for the same diagnosis		
CAGE questions, detecting alcohol abuse and dependence ⁸		
0 positive	0.1	—
1 positive	NS	—
2 positive	4.5	—
3 positive	13.3	—
4 positive	101	—
1 or more points	4.7	0.1
Compare the accuracy of the same test for the same diagnosis, in different clinical settings		
Tachypnea, detecting pneumonia in children ⁹		
All children	2.2	0.4
Disease duration <3 days	NS	NS
Disease duration 3 to 5 days	NS	NS
Disease duration ≥6 days	3.5	0.1

Diagnostic standards: for ascites, peritoneal fluid by ultrasonography; for chronic airflow obstruction, FEV₁/FEV <0.6; for myocardial infarction, cardiac enzymes or ECG, or both; for moderate and severe aortic stenosis, peak aortic flow velocity 2.6–3.5 m/s and ≥3.6 m/s, respectively; for alcohol abuse and dependence, DSM-III criteria; for pneumonia, infiltrate by chest radiograph.

Definition of findings: CAGE indicates 4 questions: have you Cut down on drinking?, have you been Annoyed by criticism?, are you Guilty about drinking?, have you ever had Eye-opener drinks?; tachypnea indicates a respiratory rate >60 breaths per minute in children <2 months old, >50 breaths per minute in children 2–12 months old, and >40 breaths per minute in children ≥1 year old.

NS, not significant, i.e., the LR's 95% confidence interval includes the value of 1.0. For the other LRs in the Table, the confidence intervals have been omitted for clarity although they all exclude the value of 1.0.

probability of ascites about 35% (LR = 6.0). Bulging flanks (LR = 2.0), flank dullness (LR = 2.0), shifting dullness (LR = 2.7), and edema (LR = 3.8) also argue for the

presence of ascites, although the increase in probability is more modest, only 15% to 25% for each finding. The most compelling argument against the diagnosis of ascites (because its LR is closest to zero) is the absence of edema, which reduces the probability of ascites about 30% (negative LR for edema = 0.2). These LRs quickly make the point that the traditional signs of ascites are not diagnostically equivalent, but that, instead, much of the diagnostic weight rests with the presence of the fluid wave (arguing for disease) and the absence of edema (arguing against it).

LRs also convey the point that positive and negative results of the same test often change probability asymmetrically. For example, in patients with chronic dyspnea, the finding of chest hyperresonance increases the probability of chronic airflow obstruction about 30% (positive LR 5.1; Table 2), yet the opposite finding, the absence of hyperresonance, has no diagnostic importance (negative LR statistically equal to 1.0).⁵

In the above examples, positive LRs increase probability and negative LRs decrease it, but this is not always the case. For example, in patients with acute chest pain, the presence of a pleuritic component to the pain argues against myocardial infarction, decreasing its probability about 30% (positive LR = 0.2). The descriptor "positive" (i.e., "positive LR") indicates the LR refers to the presence of the finding; the numerical value of the LR, being close to zero, indicates that the finding decreases probability of disease.

Clinicians may also use LRs to compare the accuracy of the same test for different definitions of disease, which usually provides important insights into the value of the test and its limitations. For example, among elderly patients with aortic flow murmurs, the finding of a late-peaking murmur argues only modestly for the diagnosis of severe aortic stenosis (LR = 3.0, increasing probability about 20%), yet the same finding is practically diagnostic for combined moderate-to-severe aortic stenosis (LR = 24.2, increasing probability about 60%).⁷ These LRs indicate that the late-peaking murmur indeed reflects greater obstruction, although it is not a pathognomonic sign of severe aortic stenosis.

LRs may also compare different levels of findings for the same diagnosis. This type of comparison is possible if the finding can be measured and placed in 3 or more levels (i.e., the finding is not simply "present" or "absent"). When levels of findings are compared, each LR becomes a "positive LR" for its own particular level and the term "negative LR" becomes meaningless. One example is the CAGE questionnaire, designed to detect alcohol abuse and dependence. A score of 0 decreases the probability of alcoholism about 45% (LR = 0.1; Table 2), whereas scores of 2, 3, or 4 increase the probability of alcoholism about 30% (LR = 4.5), 50% (LR = 13.3), or 90% (LR = 101), respectively.⁸ A CAGE score of 1 has no proven diagnostic value (i.e., its LR is statistically identical to 1.0). If the responses were instead classified into only 2 levels (e.g., a

CAGE score of 1 or more is the “positive” response and a score of 0 is “negative”), the test still discriminates between patients with and without alcoholism (positive LR 4.7, negative LR 0.1; Table 2), although these LRs obscure the point that most of the diagnostic weight of the positive response resides in scores of 3 and 4.

Finally, LRs may examine the diagnostic accuracy of the same test for the same disease but when applied to different clinical settings, a comparison that identifies in which group of patients a test is most discriminatory. For example, among children with acute respiratory complaints,⁹ the finding of tachypnea discriminates the best between those with pneumonia and those without when it is applied just to patients with symptoms lasting 6 or more days (in this group, tachypnea increases the probability of pneumonia 25% when present and decreases it 45% when absent (positive LR 3.5, negative LR 0.1; Table 2). When applied to children who are sick less than 6 days, tachypnea lacks discriminatory value (LRs not significant).

Countless published studies encompassing thousands of patients with a wide variety of disorders have investigated the diagnostic value of the patient interview, physical examination, and laboratory testing. It is very unlikely that any single clinician will personally encounter such a large number of patients, at the same time synthesizing the diagnostic impact of each part of that encounter. In a single number, LRs provide the best measure of diagnostic accuracy, and by using the simple method described in this paper, clinicians can easily take advantage of these LRs and thereby apply the lessons and

insights from published studies to their own diagnostic decisions at the bedside.

The author thanks Drs. Jan Hirschmann and Ed Boyko for their review of the manuscript and many helpful comments.

REFERENCES

1. Fagan TJ. Nomogram for Bayes' theorem. *N Engl J Med*. 1975; 293:257.
2. <http://cebmljr2.ox.ac.uk/docs/nomogram.html>. Accessed February 15, 2002.
3. Williams JW, Simel DL. Does this patient have ascites? How to divine fluid in the abdomen. *JAMA*. 1992;267:2645–8.
4. Simel DL, Halvorsen RA, Feussner JR. Quantitating bedside diagnosis: clinical evaluation of ascites. *J Gen Intern Med*. 1988; 3:423–8.
5. Badgett RG, Tanaka DJ, Hunt DK, et al. Can moderate chronic obstructive pulmonary disease be diagnosed by historical and physical findings alone? *Am J Med*. 1993;94:188–96.
6. Lee TH, Cook EF, Weisberg M, Sargent RK, Wilson C, Goldman L. Acute chest pain in the emergency room: identification and examination of low-risk patients. *Arch Intern Med*. 1985;145:65–9.
7. Aronow WS, Kronzon I. Prevalence and severity of valvular aortic stenosis determined by Doppler echocardiography and its association with echocardiographic and electrocardiographic left ventricular hypertrophy and physical signs of aortic stenosis in elderly patients. *Am J Cardiol*. 1991;67:776–7.
8. Buchsbaum DG, Buchanan RG, Centor RM, Schnoll SH, Lawton MJ. Screening for alcohol abuse using CAGE scores and likelihood ratios. *Ann Intern Med*. 1991;115:774–7.
9. Palafox M, Guiscafne H, Reyes H, Munoz O, Martinez H. Diagnostic value of tachypnoea in pneumonia defined radiologically. *Arch Dis Child*. 2000;82:41–5.