Epidemiology 3

Refining clinical diagnosis with likelihood ratios

David A Grimes, Kenneth F Schulz

Likelihood ratios can refine clinical diagnosis on the basis of signs and symptoms; however, they are underused for patients’ care. A likelihood ratio is the percentage of ill people with a given test result divided by the percentage of well individuals with the same result. Ideally, abnormal test results should be much more typical in ill individuals than in those who are well (high likelihood ratio) and normal test results should be most frequent in well people than in sick people (low likelihood ratio). Likelihood ratios near unity have little effect on decision-making; by contrast, high or low ratios can greatly shift the clinician’s estimate of the probability of disease. Likelihood ratios can be calculated not only for dichotomous (positive or negative) tests but also for tests with multiple levels of results, such as creatine kinase or ventilation-perfusion scans. When combined with an accurate clinical diagnosis, likelihood ratios from ancillary tests improve diagnostic accuracy in a synergistic manner.

Despite their usefulness in interpretation of clinical findings, laboratory tests, and imaging studies, likelihood ratios are little used. Most doctors are unfamiliar with such ratios, and few use them in practice. In a survey of 300 doctors in different specialties, only two (both internists) reported using likelihood ratios for test results.1 Since simple descriptions help clinicians to understand such ideas,2 we will try to make likelihood ratios both simple and clinically relevant.3 Our aim is to enhance clinicians’ familiarity with and use of likelihood ratios.

Some people claim that an epidemiologist sees the entire world in a 2×2 table. Indeed, if everyone could be categorised as diseased or healthy, and if a dichotomous test for that disease were universally administered, then all 6 billion of us will fit (albeit crowded) into one such table (figure 1).4 Regrettably, neither life nor tests are so simple; grey zones abound. Likelihood ratios help clinicians to navigate these large zones of clinical uncertainty.5

A likelihood ratio is simply the percentage of sick people with a given test result divided by the percentage of well individuals with the same result. A likelihood ratio, as its name implies, is the likelihood of a given test result in a person with a disease compared with the likelihood of this result in a person without the disease. Percentage and likelihood are used interchangeably here. The implications are clear: ill people should be much more likely to have an abnormal test result than healthy individuals. The size of this discrepancy has clinical importance.

Likelihood ratios for tests with two outcomes

The simple 2×2 table in the lower panel of figure 1 shows the calculation for the likelihood ratio. In this example, 15 people are sick and 12 (80%) have a true-positive test for the disease. By contrast, 85 are well but five (6%) have a false-positive test. Thus, the likelihood ratio for a positive test is simply the ratio of these two percentages (80%/6%), which is 13. Stated in another way, people with the disease are 13 times more likely to have a positive test than are those who are well. For a dichotomous test (positive or negative), this is called the positive likelihood ratio (abbreviated LR+). The flip side, the negative likelihood ratio (LR−), is calculated similarly. Three of 15 sick people (20%) have a false-negative test, whereas 80 of 85 healthy individuals (94%) have a true-negative test. So LR− is the ratio of these percentages (20%/94%), which is 0.2. Thus, a negative test is a fifth as likely in someone who is sick than in a well person. Panel 1 outlines three approaches to calculate likelihood ratios for dichotomous data.

Why bother?

Since most doctors are already familiar with terms like sensitivity and specificity,7 is learning to use likelihood ratios worth the additional effort? Likelihood ratios have several attractive features that the traditional indices of test validity do not share.8

First, not all tests have dichotomous results. Formulae for test validity do not work when results are anything other than just positive or negative. Many tests in clinical medicine have continuous results (eg, blood pressure) or multiple ordinal levels (fine-needle biopsy of breast masses).9 Collapsing multiple categories into positive and negative loses information. Likelihood ratios enable clinicians to interpret and use the full range of diagnostic test results.

Second, likelihood ratios are portable.9 By contrast, predictive values of tests are driven by the prevalence of the disease in question; even excellent tests have a poor positive predictive value when the disease is rare.10 Likelihood ratios are useful across an array of disease frequencies. While predictive values relate test characteristics to populations, likelihood ratios can be applied to a specific patient. Moreover, likelihood ratios, unlike traditional indices of validity, incorporate all four cells of a 2×2 table (panel 1).10

Third, reliance on sensitivity and specificity frequently leads to exaggeration of the benefits of tests.11 In a comparison of two obstetric tests (fetal fibronectin
measurement to predict premature birth, and uterine artery doppler wave-form analysis to predict pre-eclampsia), two-thirds of published reports overestimated the value of the tests. Use of likelihood ratios, rather than just sensitivity and specificity, might have prevented this misinterpretation.

Fourth, and most important, likelihood ratios refine clinical judgment. Application of a likelihood ratio to a working diagnosis generally changes the diagnostic probability—sometimes radically. When tests are done in sequence, the post-test odds of the first test becomes the pretest odds for the second test, and so on.

Putting likelihood ratios to work
Tests are not undertaken in a vacuum; a clinician always has an estimate (although usually not explicitly quantified) of the probability of a given disease before doing any test. According to Bayesian principles, the pretest odds of disease multiplied by the likelihood ratio gives the post-test odds of disease. For example, a pretest odds of 3/1 multiplied by a likelihood ratio of 2 would yield a post-test odds of 6/1. Unlike gamblers (or statisticians), most clinicians do not think in terms of odds—we usually use percentages. For example, a probability of 75% (75% yes/25% no) is the same as an odds of 3/1.

Although the conversion back and forth between odds and probabilities involves simple arithmetic, a widely used nomogram (figure 2, A) skirts this step altogether. A straight edge is placed on the pretest probability of disease (left column) and aligned with the likelihood ratio (middle column); the post-test probability (right column) can be read off this line. This procedure shows how much the test result has altered the pretest probability. For example, in the lower panel of figure 1, the likelihood ratio for a positive test was 13 and for a negative test, 0.2. Assume that the pretest probability of the hypothetical disease is 0.25 and that the test is positive. Placing a straight edge on a pretest probability of 0.25 and intercepting the likelihood ratio column at 13 yields a post-test probability of about 0.80, a large shift in diagnostic probability (figure 2, B). This value is close to the post-test probability of 0.81 calculated with the Bayesian formula.

Laminated copies of the nomogram are widely available. However, if working with a straight edge is unappealing, fancier methods are available. For example, a slide rule can be downloaded from the internet for calculation of post-test probabilities. The Centre for Evidence-Based Medicine in Oxford, UK, features a colourful interactive computer nomogram that uses movable arrows in lieu of a straight edge. Still, other internet programs will calculate 95% CIs around likelihood ratios for 2×2 tables. Since likelihood ratios are ratios of probabilities, we can calculate 95% CIs for them, analogous to risk ratios. Confidence intervals indicate the precision of the estimate.

Size matters
Likelihood ratios of different sizes have different clinical implications. Clinicians intuitively understand that a likelihood ratio of 1·0 is unhelpful: the percentage of sick and well people with the test result is the same. The result does not discriminate between illness and health and the pretest probability is unchanged despite the inconvenience and cost (and perhaps risk) of the test.

As with all ratios, likelihood ratios start at unity and extend down to zero and up to infinity. Hence, the further the likelihood ratio is from 1·0, the greater its

Panel 1: Calculation of likelihood ratios for dichotomous results
If sensitivity and specificity have already been determined, then
LR+ is sensitivity/(1−specificity)
LR− is (1−sensitivity)/specificity
If raw numbers for the 2×2 table are available, then
LR+ is (a/(a+c))/(b/(b+d))
LR− is (c/(a+c))/(d/(b+d))
If mathematical formulas are unappealing, then
LR+ is the true-positive percent divided by the false-positive percent
LR− is the false-negative percent divided by the true-negative percent
Table 1: Likelihood ratios for white-blood-cell count in diagnosing appendicitis

<table>
<thead>
<tr>
<th>n (%) with appendicitis</th>
<th>n (%) without appendicitis</th>
<th>% with appendicitis/ % without appendicitis</th>
<th>Likelihood ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>=7×10^9 cells per L</td>
<td>1 (2%)</td>
<td>30 (23%)</td>
<td>0·10</td>
</tr>
<tr>
<td>7.9×10^9 cells per L</td>
<td>9 (15%)</td>
<td>47 (29%)</td>
<td>0·52</td>
</tr>
<tr>
<td>9.11×10^9 cells per L</td>
<td>4 (7%)</td>
<td>35 (24%)</td>
<td>0·29</td>
</tr>
<tr>
<td>11-13×10^9 cells per L</td>
<td>22 (37%)</td>
<td>19 (13%)</td>
<td>3·7</td>
</tr>
<tr>
<td>13-15×10^9 cells per L</td>
<td>6 (10%)</td>
<td>9 (6%)</td>
<td>1·7</td>
</tr>
<tr>
<td>15-17×10^9 cells per L</td>
<td>8 (14%)</td>
<td>7 (5%)</td>
<td>2·8</td>
</tr>
<tr>
<td>17-19×10^9 cells per L</td>
<td>4 (7%)</td>
<td>3 (2%)</td>
<td>3·5</td>
</tr>
<tr>
<td>≥19×10^9 cells per L</td>
<td>5 (8%)</td>
<td>0</td>
<td>8·0</td>
</tr>
</tbody>
</table>

Adapted from reference 19 with permission of the American College of Emergency Physicians.

Table 1: Likelihood ratios for white-blood-cell count in diagnosing appendicitis

The likelihood ratio for a count of 7×10^9 cells per L, 2% is the numerator (those with appendicitis) and 21% the denominator (those without appendicitis); the likelihood ratio is 2%/21%, or 0·1. This same calculation is done for every level of white-blood-cell count; for the highest values, the calculation cannot be done because the denominator is zero. Likelihood ratios vary from 0·1 to infinity, with a trend towards higher ratios with higher white-blood-cell counts.

But will these likelihood ratios change practice? Will they either lower the diagnostic probability enough to send a patient home from the emergency department or raise it sufficiently to head to the operating theatre? Most patients (82%) had white-blood-cell counts between 7 and 19×10^9 cells per L; the resultant likelihood ratios ranged from 0·52 to 3·5, which have little effect on probability. Stated alternatively, in four-fifths of patients being assessed for appendicitis, the white-blood-cell count was not helpful in reaching a diagnosis.19 Only extreme values would shift the probability much. Consider a 28-year-old man with a 20% pretest probability of pulmonary embolism. He has a ventilation-perfusion scan interpreted as normal, which has a likelihood ratio of 0·1.20 If we place a straight-edge at 20% in the left column of the nomogram and align it with 0·1 in the middle column, the right column indicates a post-test probability around 2% (figure 2, C).

Prostate-specific antigen screening for prostate cancer provides another example of multiple likelihood ratios.20
In a community-based study of 2620 men age 40 years or older, investigators did prostate-specific antigen testing and used prostate biopsy as the diagnostic gold standard. With the standard cutoff of 4 μg/L, the likelihood ratio for a positive test was 1.3 (95% CI 1.2–1.3) and for a negative test 0.4 (0.4–0.5)—not much help clinically. However, when broken down by concentrations of prostate-specific antigen, results are more useful (table 2). The lowest value (<2 μg/L) had a likelihood ratio of 0.3 and the highest (≥20 μg/L) a ratio of 6.3. These likelihood ratios would yield moderate changes in the pretest probability of cancer.

A useful mnemonic
Regrettably, nomograms and computers are usually not available at the bedside. Hence, a mnemonic suggested by McGee for simplifying the use of likelihood ratios has strong appeal. He notes that for pretest probabilities between 10% and 90% (the usual situation), the change in probability from a test or clinical finding is approximated by a constant. The clinician needs to remember only three benchmark likelihood ratios: 2, 5, and 10 (table 3). These correspond to the first three multiples of 15%; a likelihood ratio of 2 increases the probability by about 15%, 5 by 30%, and 10 by 45%. For example, with a pretest probability of 40% and a likelihood ratio of 2, the post-test probability is 40%+15%=55% (quite close to the 57% when calculated by formula). For likelihood ratios less than 1, the rule works in the opposite direction. The reciprocal of 2 is 0.5; that of 5 is 0.2, and that of 10 is 0.1. A likelihood ratio of 0.5 would reduce the pretest probability by about 15% while a ratio of 0.1 would drop it by about 45 absolute percentage points.

The importance of accurate pretest probability
The medical history and physical examination remain fundamentally important. Indeed, a precise assessment of the chance of disease can be far more important than the likelihood ratios stemming from expensive, sometimes invasive tests. For some diseases, such as Alzheimer’s dementia and sinusitis, clinical findings yield a highly accurate diagnosis. For other diseases, clinicians lack information about the predictive value of signs and symptoms; here they must rely on epidemiological data, education, and clinical acumen. For example, if additional patient history revised a pretest probability of coronary disease from 75% to less than 5%, this change would affect the post-test probability of disease more than would a stress test with positive and negative likelihood ratios of 3 and 0.5, respectively. Although clinical diagnosis might not necessarily be more accurate than ancillary testing, its precision has a striking effect on the interpretation of any test results that follow. An accurate pretest probability and subsequent testing can greatly improve clinical diagnosis.

Diagnostic thresholds
Tests should only be used when they will affect management. If a clinician’s pretest probability of disease securely rules in or out a diagnosis, further testing is unwarranted. More testing should be considered only in the murky middle zone of clinical uncertainty (figure 3). The location of these decision thresholds (A and B) along the continuum of diagnostic certainty needs to be determined before testing is done. Probabilities lower than point A effectively exclude the
diagnosis in question. Hence, point A becomes the testing threshold: pretest probabilities greater than A but lower than B could benefit from further testing. Point B is the treatment threshold: probabilities greater than this point justify beginning treatment without further delay.

The locations of these decision thresholds (A and B) should be tailored to the specific patient. Using the nomogram (figure 2, A), a clinician can estimate how high or low a likelihood ratio would have to be to shift the pretest probability below A (exclude the diagnosis) or above B (begin treatment).25 A clinician can consult published likelihood ratios for tests to find the corresponding test values.22,24 If no test result would achieve this shift in probability, the test should not be done—a fundamentally important point.

Limitations of likelihood ratios
The effect of likelihood ratios on pretest probabilities is not linear. A likelihood ratio of 100 does not increase the pretest probability ten times more than does a ratio of 10, as figure 2, D shows.

For tests with several categories of results, extreme test values yield imprecise likelihood ratios. Few patients having values that are either very high or low result in little precision. Small changes in the numbers of patients in these cells can produce very different likelihood ratios. Stated alternatively, imprecision in likelihood ratios is greatest at the top and bottom of test-result distributions.25 Combining continuous categories at the extremes of the test-result distribution provides larger numbers and more precision—ie, narrower confidence intervals.26

Conversely, many test results will fall towards the centre of the distribution. Here, likelihood ratios are closer to 1 and thus help little. The big payoffs stem from high or low likelihood ratios.26 An additional problem is that pretest probabilities developed in tertiary-care settings might not be applicable because of differences in patient populations.26 Panel 2 provides some guidelines for ordering tests on the basis of pretest probabilities.

Uses for likelihood ratios
Likelihood ratios have a broad array of clinical applications, including symptoms, physical examinations, laboratory tests, imaging procedures, and scoring systems (table 4). Several resources have compiled reported likelihood ratios, including a handbook20 that contains more than 140. Another publication includes ratios for both diagnostic tests and clinical findings.22 Building on an accurate pretest probability of disease, likelihood ratios from ancillary tests can refine clinical judgment—often in important ways.

Conflict of interest statement
We declare that we have no conflict of interest.

Acknowledgments
We thank Willard Cates and David L Sackett for their helpful comments on an earlier version of this report.

References
Series