Epidemiology 3

Refining clinical diagnosis with likelihood ratios

Lancet 2005; 365: 1500–05

David A Grimes, Kenneth F Schulz

Family Health International, PO Box 13950, Research Triangle Park, NC 27709, USA
(D A Grimes MD, K F Schulz PhD)

Correspondence to:
Dr David A Grimes
dgrimes@fhi.org

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,[2] 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.[4]

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).[6–8] 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.[4] 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.[9] 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,[12] a widely used nomogram[13] (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.[9] 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.[14] 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.[15] Still, other internet programs will calculate 95% CIs around likelihood ratios for 2×2 tables.[16] Since likelihood ratios are ratios of probabilities, we can calculate 95% CIs for them, analogous to risk ratios.[17] Confidence intervals indicate the precision of the estimate.

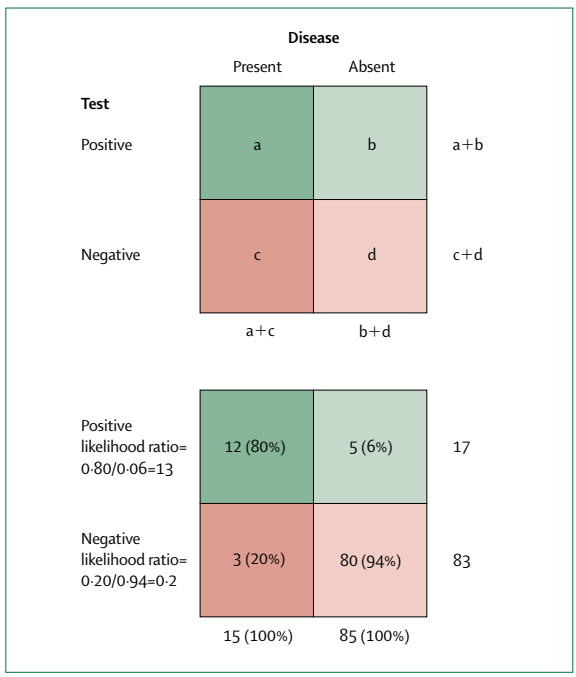

Figure 1: **2×2 tables**
Upper panel shows distribution of population by disease status and dichotomous test result. Lower panel shows hypothetical distribution of 100 people by disease status and dichotomous test result.

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

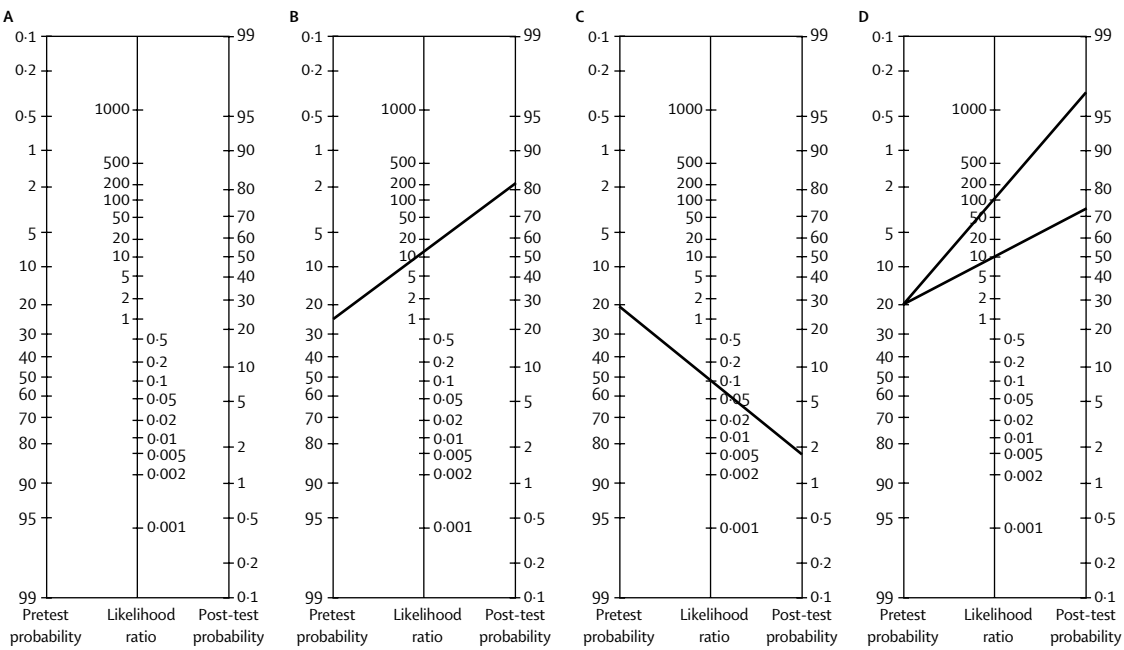

Figure 2: **Nomograms for probabilities and likelihood ratios**[13]
(A) Nomogram reprinted from reference 13 with permission of the Massachusetts Medical Association. (B) Straight edge applied for pretest probability of 0·25 and likelihood ratio of 13. (C) Straight edge applied for pretest probability of 0·20 and likelihood ratio of 0·1. (D) Effect of likelihood ratios of 10 and 100 on pretest probability of 0·2.

effect is on the probability of disease. Likelihood ratios from 2 to 5 yield small increases in the post-test probability of disease, from 5 to 10 moderate increases, and above 10 large increases. For ratios less than unity, the smaller the likelihood ratio, the greater the decrease in probability.[18]

Likelihood ratios for tests with multiple outcomes

Calculation of likelihood ratios for tests with more than two outcomes is similar to the calculation for dichotomous outcomes; a separate likelihood ratio is simply calculated for every level of test result. In table 1, white-blood-cell counts are shown for 59 patients with appendicitis and 145 without the diagnosis. To calculate 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.

	n (%) with appendicitis	n (%) without appendicitis	% with appendicitis/ % without appendicitis	Likelihood ratio
≤7×10^9 cells per L	1 (2%)	30 (21%)	2/21	0·10
7–9×10^9 cells per L	9 (15%)	42 (29%)	15/29	0·52
9–11×10^9 cells per L	4 (7%)	35 (24%)	7/24	0·29
11–13×10^9 cells per L	22 (37%)	19 (13%)	37/13	2·8
13–15×10^9 cells per L	6 (10%)	9 (6%)	10/6	1·7
15–17×10^9 cells per L	8 (14%)	7 (5%)	14/5	2·8
17–19×10^9 cells per L	4 (7%)	3 (2%)	7/2	3·5
≥19×10^9 cells per L	5 (8%)	0	8/0	∞
Total	59 (100%)	145 (100%)		

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**

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.[12] 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]

	Number of men tested	Likelihood ratio (95% CI)
<2 μg/L	378	0·3 (0·2–0·3)
≥2 to 4 μg/L	313	0·7 (0·6–0·9)
>4 to 10 μg/L	1302	1·0 (0·9–1·0)
>10 to 20 μg/L	421	1·5 (1·2–1·8)
>20 μg/L	206	6·3 (4·6–8·7)

Adapted from reference 20 with permission of BioMed Central.

Table 2: **Likelihood ratios for prostate-specific antigen in diagnosing prostate cancer**

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.[20] 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.[21] 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.

	Approximate change in probability (%)
Likelihood ratios between 0 and 1 reduce the probability of disease	
0·1	−45
0·2	−30
0·3	−25
0·4	−20
0·5	−15
1·0	0
Likelihood ratios greater than 1 increase the probability of disease	
2	+15
3	+20
4	+25
5	+30
6	+35
7	
8	+40
9	
10	+45

Reproduced from reference 21 with permission of Blackwell Publishing.

Table 3: **Likelihood ratios and bedside estimates**

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.[22] 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.[22] 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[23] (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

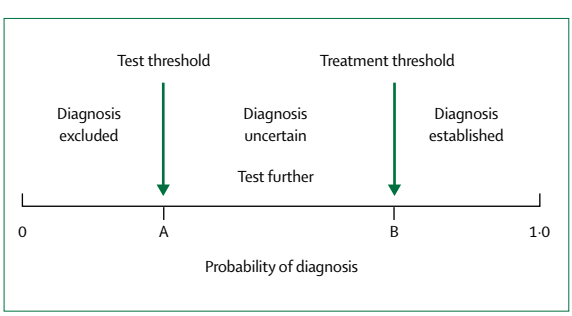

Figure 3: **Thresholds for testing and treating, as a function of probability of diagnosis**

	Disease or outcome
Physical examination	
Anisocoria	Cause of coma[29]
Signs or symptoms	Childhood tinea capitis[30]
Physical examination	Posterior pelvic ring injuries in trauma patients[31]
Clinical findings	Acute bacterial sinusitis[32]
Clinical findings	Symptomatic sacroiliac joints[33]
Ottawa ankle rules	Fractures of the ankle and midfoot[34]
Laboratory tests	
Urinalysis	Urinary-tract infection in children[35]
Throat swab	Chronic tonsillitis[36]
Chlamydia trachomatis antibody testing	Tubal factor infertility[37]
Prostate-specific antigen	Prostate cancer[38]
Blood culture	Postoperative mediastinitis[39]
Imaging studies	
Transvaginal ultrasound examination	Ovarian endometrioma[40]
Screening mammography	Breast cancer[41]
Scoring system	
Modified organ system failure score	Discharge outcome from intensive care unit[42]
Ongoing abuse screen	Intimate partner violence[43]
Other	
Surname	Chinese ancestry[44]

Table 4: **Examples of likelihood ratio applications**

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).[18] 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.

Panel 2: Tips on testing

- Clinicians should be wary of ordering tests when the pretest probability of disease is high or low.[27] Tests are unlikely to alter disease probability and will only confuse the situation: unexpected results will usually be false-positives or false-negatives.
- Tests are most useful when the pretest probability is 50%.[27] Numerical changes in the post-test column of the nomogram (figure 2) are greater when the starting point in the pretest column is at 50% than elsewhere.
- The higher the pretest probability of disease, the higher will be the post-test probability, no matter what the test result. For example, three times a high probability will be larger than three times a low one.[27]
- LR+ greater than 10 means that a positive test is good at ruling in a diagnosis.[28]
- A likelihood ratio negative less than 0·1 means that a negative test is good at ruling out a diagnosis.[28]
- When using tests in sequence, the post-test probability of the first test becomes the pretest probability for the next one.[28] Tests can build on each other in sequence.

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 handbook[24] 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

1 Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med* 1998; **104:** 374–80.

2 Steurer J, Fischer JE, Bachmann LM, Koller M, ter Riet G. Communicating accuracy of tests to general practitioners: a controlled study. *BMJ* 2002; **324:** 824–26.

3 Feinstein AR. Clinical biostatistics: XXXIX—the haze of Bayes, the aerial palaces of decision analysis, and the computerized Ouija board. *Clin Pharmacol Ther* 1977; **21:** 482–96.

4 Grimes DA, Schulz KF. Uses and abuses of screening tests. *Lancet* 2002; **359:** 881–84.

5 Naylor CD. Grey zones of clinical practice: some limits to evidence-based medicine. *Lancet* 1995; **345:** 840–42.

6 Choi BC. Slopes of a receiver operating characteristic curve and likelihood ratios for a diagnostic test. *Am J Epidemiol* 1998; **148:** 1127–32.

7 Giard RW, Hermans J. Interpretation of diagnostic cytology with likelihood ratios. *Arch Pathol Lab Med* 1990; **114:** 852–54.

8 Lee WC. Selecting diagnostic tests for ruling out or ruling in disease: the use of the Kullback-Leibler distance. *Int J Epidemiol* 1999; **28:** 521–25.

9 Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology: a basic science for clinical medicine, 2nd edn. Boston: Little, Brown and Company, 1991.

10 Chien PF, Khan KS. Evaluation of a clinical test: II—assessment of validity. *BJOG* 2001; **108:** 568–72.

11 Khan KS, Khan SF, Nwosu CR, Arnott N, Chien PF. Misleading authors' inferences in obstetric diagnostic test literature. *Am J Obstet Gynecol* 1999; **181:** 112–15.

12 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:** 703–07.

13 Fagan TJ. Letter: nomogram for Bayes theorem. *N Engl J Med* 1975; **293:** 257.

14 Children's Mercy Hospital. Likelihood ratio slide rule. http://www.childrens-mercy.org/stats/sliderule.asp (accessed April 4, 2005).

15 Centre for Evidence-Based Medicine. Nomogram for likelihood ratios. http://www.cebm.net/nomogram.asp (accessed April 4, 2005).

16 Herbert R. Confidence interval calculator (version 4, November, 2002). http://www.pedro.fhs.usyd.edu.au/Utilities/CIcalculator.xls (accessed April 4, 2005).

17 Altman DG. Diagnostic tests. In: Altman DG, Machin D, Bryant TN, Gardner MJ, eds. Statistics with confidence, 2nd edn. London: BMJ Books, 2000: 105–19.

18 Hayden SR, Brown MD. Likelihood ratio: a powerful tool for incorporating the results of a diagnostic test into clinical decision making. *Ann Emerg Med* 1999; **33:** 575–80.

19 Snyder BK, Hayden SR. Accuracy of leukocyte count in the diagnosis of acute appendicitis. *Ann Emerg Med* 1999; **33:** 565–74.

20 Hoffman RM, Gilliland FD, Adams-Cameron M, Hunt WC, Key CR. Prostate-specific antigen testing accuracy in community practice. *BMC Fam Pract* 2002; **3:** 19.

21 McGee S. Simplifying likelihood ratios. *J Gen Intern Med* 2002; **17:** 646–49.

22 Halkin A, Reichman J, Schwaber M, Paltiel O, Brezis M. Likelihood ratios: getting diagnostic testing into perspective. *QJM* 1998; **91:** 247–58.

23 Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med* 1980; **302:** 1109–17.

24 Goldenberg K, Verdain Barnes H, Redding MM. Diagnostic testing handbook for clinical decision making. Chicago: Yearbook Medical Publishers, 1989.

25 Dujardin B, Van den Ende J, Van Gompel A, Unger JP, Van der Stuyft P. Likelihood ratios: a real improvement for clinical decision making? *Eur J Epidemiol* 1994; **10:** 29–36.

26 Sonis J. How to use and interpret interval likelihood ratios. *Fam Med* 1999; **31:** 432–37.

27 Sharma S. The likelihood ratio and ophthalmology: a review of how to critically appraise diagnostic studies. *Can J Ophthalmol* 1997; **32:** 475–78.

28 Davidson M. The interpretation of diagnostic tests: a primer for physiotherapists. *Aust J Physiother* 2002; **48:** 227–32.

29 Tokuda Y, Nakazato N, Stein GH. Pupillary evaluation for differential diagnosis of coma. *Postgrad Med J* 2003; **79:** 49–51.

30 Hubbard TW. The predictive value of symptoms in diagnosing childhood tinea capitis. *Arch Pediatr Adolesc Med* 1999; **153:** 1150–53.

31 McCormick JP, Morgan SJ, Smith WR. Clinical effectiveness of the physical examination in diagnosis of posterior pelvic ring injuries. *J Orthop Trauma* 2003; **17:** 257–61.

32 DeAlleaume L, Parker S, Reider JM. What findings distinguish acute bacterial sinusitis? *J Fam Pract* 2003; **52:** 563–65.

33 Laslett M, Young SB, Aprill CN, McDonald B. Diagnosing painful sacroiliac joints: a validity study of a McKenzie evaluation and sacroiliac provocation tests. *Aust J Physiother* 2003; **49:** 89–97.

34 Bachmann LM, Kolb E, Koller MT, Steurer J, ter Riet G. Accuracy of Ottawa ankle rules to exclude fractures of the ankle and mid-foot: systematic review. *BMJ* 2003; **326:** 417.

35 Bachur R, Harper MB. Reliability of the urinalysis for predicting urinary tract infections in young febrile children. *Arch Pediatr Adolesc Med* 2001; **155:** 60–65.

36 Kurien M, Stanis A, Job A, Brahmadathan, Thomas K. Throat swab in the chronic tonsillitis: how reliable and valid is it? *Singapore Med J* 2000; **41:** 324–26.

37 Dabekausen YA, Evers JL, Land JA, Stals FS. Chlamydia trachomatis antibody testing is more accurate than hysterosalpingography in predicting tubal factor infertility. *Fertil Steril* 1994; **61:** 833–37.

38 Meigs JB, Barry MJ, Oesterling JE, Jacobsen SJ. Interpreting results of prostate-specific antigen testing for early detection of prostate cancer. *J Gen Intern Med* 1996; **11:** 505–12.

39 Fowler VG Jr, Kaye KS, Simel DL, et al. *Staphylococcus aureus* bacteremia after median sternotomy: clinical utility of blood culture results in the identification of postoperative mediastinitis. *Circulation* 2003; **108:** 73–78.

40 Moore J, Copley S, Morris J, Lindsell D, Golding S, Kennedy S. A systematic review of the accuracy of ultrasound in the diagnosis of endometriosis. *Ultrasound Obstet Gynecol* 2002; **20:** 630–34.

41 Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Likelihood ratios for modern screening mammography; risk of breast cancer based on age and mammographic interpretation. *JAMA* 1996; **276:** 39–43.

42 Eapen CE, Thomas K, Cherian AM, Jeyaseelan L, Mathai D, John G. Predictors of mortality in a medical intensive care unit. *Natl Med J India* 1997; **10:** 270–72.

43 Weiss SJ, Ernst AA, Cham E, Nick TG. Development of a screen for ongoing intimate partner violence. *Violence Vict* 2003; **18:** 131–41.

44 Choi BC, Hanley AJ, Holowaty EJ, Dale D. Use of surnames to identify individuals of Chinese ancestry. *Am J Epidemiol* 1993; **138:** 723–34.

*N Horgan, E McLoone, B Lannigan, I Flitcroft
noelhorg@hotmail.com

Department of Ophthalmology, Children's University Hospital, Temple Street, Dublin 1, Ireland

1 Wagoner MD. Chemical injuries of the eye: current concepts in pathophysiology and therapy. *Surv Ophthalmol* 1997; **41:** 275–313.
2 Friend J, Kiorpes T, Thoft RA. Conjunctival goblet cell frequency after alkali injury is not accurately reflected by aqueous tear mucin content. *Invest Ophthalmol Vis Sci* 1983; **24:** 612–18.
3 Brodowsky SC, McCarty CA, Snibson G, et al. Management of alkali burns—an 11 year retrospective review. *Ophthalmol* 2000; **107:** 1829–35.

The trouble with likelihood ratios

David Grimes and Kenneth Schulz (Apr 23, p 1500)[1] state that likelihood ratios are underused in patients' care. For years we have tried in vain to introduce them in clinical teaching in four continents, and we hoped desperately to see diagnostic research published as likelihood ratios. We have analysed some possible explanations for this defeat.

The first is the complex chain of calculations involved. Clinicians should transform probabilities into odds, multiply by a series of likelihood ratios, and finally reconvert odds to probabilities. Simplifying aids such as the Fagan nomogram are rarely used, should be done for every test, and need published likelihood ratios.

The second problem is the absence of an appropriate language for clinical logic. Instead of indicating what it means for clinicians, the word "likelihood ratio" states where it comes from. Let us give a corollary: should we tell a violinist he or she should play the "dominant ratio" (3/2) of the A? No, we ask him or her to play a "fifth", a logarithmic "metalanguage" traduction. A positive likelihood ratio means "confirming power" to the clinician, so why not call it that?[2] Never, in 20 years of teaching clinical logic, have we found a clinician who used the word "positive likelihood ratio". Furthermore, there is no word for "odds" in French, Italian, Spanish, Kinyarwanda, or Lao, to name but examples. So some have to start difficult calculations with a notion that does not even exist in their mind.

A third and fundamental problem is the counterintuitive scale of likelihood ratios. Why is a test with a likelihood ratio of 100 not 10 times more powerful than a test with a likelihood ratio of 10? Why is the likelihood ratio given in strange numbers such as 0·01? How to compare the excluding power of 0·03 with a confirming power of 33? And why do confidence intervals widen for high and very low likelihood ratios? Mathematically speaking, likelihood ratios and odds both have a skewed, exponential distribution. The great mathematician Turing proposed that likelihood ratios be represented in a logarithmic way, and be grouped in classes.[3]

Grimes and Schulz state that tests are most useful when used around 50% probability. This is the last (and not least) clue to the lack of success of the actual model. When an HIV ELISA test alters your probability from 0·1 to 10, is the diagnostic benefit less than if it were changed from 10 to 90? The same holds for a biopsy that pushes the probability from 98 to 99·99. The probability scale for clinicians is symmetrically skewed in both directions. In primary care, elective surgery, and oncology, clinicians work at the extremes of the probability scale.

The time has come to apply the Feynman-Tufte principle to clinical logic, to offer a visual representation of Bayesian logic.[4] Earlier attempts with natural logarithmics do not allow a simple graphic.[5] We teach a scale of $\log_{10}$ odds from −4 to +4, indicating the corresponding probabilities, and adding rounded $\log_{10}$ likelihood ratios to the pretest probability. In doing so, clinicians can apply Bayesian logic without formal calculations (figure).

We declare that we have no conflict of interest.

*Jef Van den Ende, Juan Moreira, Paulin Basinga, Zeno Bisoffi
jvdende@itg.be

*Department of Clinical Sciences, Institute of Tropical Medicine, Antwerp, Belgium (JVdE); Centro de Epidemiología Comunitaria y Medicina Tropical, Esmeraldas, Ecuador (JM); National University of Rwanda, Butare, Rwanda (PB); and Centro per le Malattie Tropicali, Negrar, Verona, Italy (ZB)

1 Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet* 2005; **365:** 1500–05.
2 Van Puymbroeck H, Remmen R, Denekens J, Scherpbier A, Bisoffi Z, Van den Ende J. Teaching problem solving and decision making in undergraduate medical education: an instructional strategy. *Med Teach* 2003; **25:** 547–50.
3 Good IJ. AM Turing's statistical work in world war II. Studies in the history of probability and statistics XXXVII. *Biometrika* 1979; **66:** 393–96.
4 Shermer M. The Feynman-Tufte principle. *Sci Am* 2005; **4:** 21.
5 Glass RD. Diagnosis: a brief introduction, 1st edn. Melbourne: Oxford University Press, 1996.

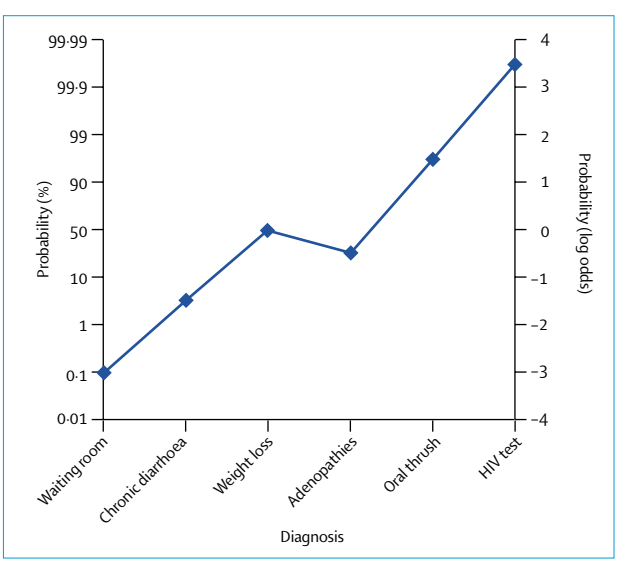

Figure: **Evolution of probability in a patient suspected of having HIV infection, following consecutive diagnostic steps**
At the left hand side, probability is shown in percentages, at the right hand side in log odds. Log_{10} likelihood ratios, rounded to half the unit, are added to the pretest probability.

Department of Error

Kapp C. Hamilton Naki. Lancet *2005;* **366:** 22—In this Obituary (July 2), Clare Kapp described how Hamilton Naki took part in the first successful heart transplant with Christiaan Barnard. Naki was not present during this operation. The surgeons who removed the donor's heart were Marius Barnard and Terry O'Donovan. Naki was a skilled laboratory assistant employed by the University of Cape Town. He did not operate on human beings, nor did he ever work within the Groote Schuur Hospital, or its operating theatres. Naki's role was restricted to work on animals and he assisted Christiaan Barnard in the research effort that preceded the first human heart transplant.